

A Lipid Metabolism-Associated Six-Gene Signature for Survival Stratification in Lung Squamous Cell Carcinoma with Experimental Validation of Selected Genes

Shan Muhammad

nimbus14@hotmail.com

Changsha Medical University

JiaMing Xie

National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College

Daniel Muteb Muyey

Changsha Medical University

Mamona Bilal

Second Affiliated Hospital of Chongqing Medical University

Yang Hai

Public Health College of Harbin Medical University

Jie He

Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College

Research Article

Keywords: Lung squamous cell carcinoma, lipid metabolism, prognostic signature, bioinformatics, tumor microenvironment, ALOX15B, PPP2R2C, translational oncology

Posted Date: September 29th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9717267/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

1 **Title: A Lipid Metabolism-Associated Six-Gene Signature for**
2 **Survival Stratification in Lung Squamous Cell Carcinoma with**
3 **Experimental Validation of Selected Genes**

4 Shan Muhammad^{1,2,3,4, #*}, JiaMing Xie^{2,3#}, Daniel Muteb Muyey^{1#}, Mamona Bilal^{5#},
5 Yang Hai⁶, Jie He^{2,3*}

6 ¹ Department of International Medical Education, Changsha Medical University,
7 Wangcheng District, Changsha, 410219, China

8 ² Department of Thoracic Surgery, National Cancer Center/National Clinical Research
9 Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking
10 Union Medical College, Beijing, 100021, China

11 ³ Laboratory of Translational Medicine, National Cancer Center/National Clinical
12 Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and
13 Peking Union Medical College, Beijing 100021, China

14 ⁴ Department of Colorectal Surgery, National Cancer Center/National Clinical Research
15 Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences & Peking
16 Union Medical College, Beijing, 100021, China

17 ⁵ Department of Obstetrics and Gynaecology, The Second Affiliated Hospital of
18 Chongqing Medical University, Chongqing, 400016, China.

19 ⁶ Department of Children's and Adolescent Health, Public Health College of Harbin
20 Medical University, 157 Baojian Road, Harbin, 150081, China.

21 ***Correspondence:** Shan Muhammad

22 nimbus14@hotmail.com

Tel: +86-13520863407

Jie He

hejie.cams@gmail.com

Tel: +86-13936443107

[#]These authors contributed equally to this work and shared first authorship

Abstract

Background

Lung squamous cell carcinoma remains a clinically important subtype of non-small cell lung cancer with limited biologically grounded biomarkers for risk stratification. Because lipid metabolic reprogramming contributes to tumor progression and microenvironmental remodeling, we performed a retrospective multi-cohort transcriptomic analysis to develop and evaluate a lipid metabolism-associated prognostic signature in lung squamous cell carcinoma, explore its biological and immune-related features, and provide experimental validation for selected genes within the model.

Results

A total of 1,064 lipid metabolism-associated genes were screened, of which 112 were differentially expressed in the TCGA lung squamous cell carcinoma cohort. These genes were mainly enriched in arachidonic acid metabolism, ether lipid metabolism, glycerophospholipid metabolism, and related metabolic pathways. A six-gene prognostic signature comprising *ALOX15B*, *CYP24A1*, *PPP2R2C*, *PTGIS*, *SPTSSB*, and *UGT2B17* was constructed in the TCGA training cohort and further evaluated in a TCGA internal test set and six external Gene Expression Omnibus cohorts. The signature stratified

patients into high- and low-risk groups with significantly different overall survival and remained independently associated with prognosis in multivariable analyses. In GSE74777, the signature was also associated with recurrence-free survival, with high-risk patients showing poorer recurrence-free survival than low-risk patients. Risk-group comparisons indicated distinct pathway activity and immune-cell composition, and exploratory in silico analyses suggested differences in selected immunotherapy-response-related features. Connectivity Map analysis identified putative compounds for future investigation. To strengthen biological plausibility, selected genes within the signature were further examined in lung squamous cell carcinoma cell lines. Baseline quantitative real-time PCR and western blotting supported differential expression of selected genes, siRNA-mediated suppression of *ALOX15B* and *PPP2R2C* was confirmed, and functional assays provided additional support for a potential role of *ALOX15B* in clonogenic growth and RSL3-related viability responses.

Conclusions

This study defines a lipid metabolism-associated six-gene prognostic framework for lung squamous cell carcinoma that is supported by multi-cohort retrospective evaluation and selected experimental validation. The findings link survival stratification to biologically relevant metabolic and immune-related features and provide a basis for further mechanistic and prospective translational investigation.

Keywords: Lung squamous cell carcinoma; lipid metabolism; prognostic signature; bioinformatics; tumor microenvironment; *ALOX15B*; *PPP2R2C*; translational oncology

Introduction

Lung cancer remains the leading cause of cancer-related mortality worldwide, and non-small cell lung cancer (NSCLC) accounts for the great majority of cases (1). Within NSCLC, lung squamous cell carcinoma (LUSC) remains a clinically important subtype because it is biologically distinct from lung adenocarcinoma and still has fewer established biomarker-guided treatment options and fewer actionable therapeutic avenues than other thoracic malignancies (2). LUSC is also more strongly associated with tobacco exposure and remains more common in men, underscoring its distinct epidemiologic and biological profile within thoracic oncology (2-4). Although advances in screening, perioperative management, systemic therapy, and molecular characterization have improved outcomes for selected patients, prognosis for LUSC remains heterogeneous, and clinically meaningful stratification beyond conventional clinicopathologic parameters remains an unmet need (1, 2, 5). In this setting, biomarkers that are not only statistically associated with outcome but also anchored in plausible tumor biology are of particular translational interest.

Metabolic reprogramming is a core feature of malignant progression, and lipid metabolism has emerged as a particularly relevant component of this process (6-8). Beyond serving as an energy source, lipids contribute to membrane biogenesis, signaling, redox homeostasis, stress adaptation, and survival under nutrient- and oxygen-limited conditions (7, 9, 10). Lipid metabolic remodeling can also shape the tumor microenvironment by influencing inflammatory mediators, immune-cell state, metastatic competence, and the balance between anabolic growth and oxidative vulnerability (7, 8, 11). In thoracic malignancies, growing interest in metabolic liabilities has further highlighted lipid pathways as biologically meaningful determinants of tumor behavior

and as a source of potential biomarkers for disease stratification and hypothesis generation (12). Taken together, these observations provide a strong rationale for examining lipid metabolism not simply as a generic hallmark of cancer, but as a disease-relevant axis with implications for tumor aggressiveness and microenvironmental context in LUSC (12-14).

In LUSC specifically, metabolic dysregulation has increasingly been linked to phenotype, prognosis, and immune context, but the translational biomarker landscape remains fragmented. Prior studies have shown that lipid-related genes may be associated with LUSC progression, tumor immunity, and survival, and metabolism-oriented prognostic models have already been reported for this disease (15, 16). These studies collectively support the biological relevance of lipid pathways in LUSC, but they also illustrate a recurring challenge in the field: many transcriptome-derived signatures are retrospective, vary considerably in composition and interpretability, and provide uneven levels of biological support. As a result, the key question is no longer whether metabolism-associated signatures can be generated for LUSC, but whether they can be framed in a way that is biologically grounded, clinically interpretable, and supported by more than computational association alone (16, 17). In this context, a more meaningful contribution is not simply the generation of another prognostic signature, but the development of a lipid metabolism-centered framework that is evaluated across multiple cohorts and supported by selected experimental evidence.

This gap matters because prognostic biomarker studies are most informative when they do more than separate survival curves. In translational oncology, the value of such work lies in linking human tumor-derived molecular patterns to biologically plausible

processes that can guide future mechanistic study and improve risk stratification without overstating immediate clinical implementation. We therefore conducted an integrated multi-cohort analysis of LUSC transcriptomic data to identify lipid metabolism-associated differentially expressed genes, construct a prognostic signature for survival stratification, and characterize the biological and immune-related features associated with that signature. In addition, to strengthen the biological relevance of the computational findings, we performed experimental validation of selected genes within the model using LUSC cell lines. Through this combined bioinformatic and experimental approach, we sought to develop a more biologically supported lipid metabolism-associated prognostic framework for LUSC.

Materials and Methods

Study design and overall workflow

This study was designed as a retrospective, multi-cohort transcriptomic and clinical bioinformatics investigation of lung squamous cell carcinoma (LUSC), integrated with selected in vitro validation experiments. Publicly available LUSC gene-expression datasets with survival information were used to identify lipid metabolism-associated differentially expressed genes, develop a prognostic signature, and evaluate its association with biological pathways and immune-related features. The computational findings were subsequently complemented by experimental validation of selected genes within the final model in LUSC cell lines. The overall workflow of the study is summarized in **Figure 1** and comprised four major stages: (i) collection and preprocessing of discovery and validation cohorts; (ii) construction and retrospective

evaluation of a lipid metabolism-associated prognostic signature; (iii) exploratory in silico analyses of pathway enrichment, immune infiltration, immunotherapy-response-related biomarkers, and candidate compounds; and (iv) experimental validation of selected genes, with functional follow-up centered on *ALOX15B* and *PPP2R2C*.

Data acquisition and cohort selection

RNA sequencing (RNA-seq) data and corresponding clinical annotations for LUSC were obtained from The Cancer Genome Atlas (TCGA) portal, including 502 primary tumor samples and 49 normal lung tissue samples. TCGA-LUSC data were downloaded on July 3, 2022 (<https://portal.gdc.cancer.gov/repository>). Only primary tumor samples (sample type code 01A) and matched normal control samples (sample type code 11A) were retained for differential-expression analysis. Patients lacking complete overall survival (OS) information were excluded from prognostic modeling, leaving 474 TCGA-LUSC cases for downstream survival analyses. Baseline clinicopathological characteristics of the TCGA-LUSC cohort are summarized in **Table 1**.

External validation cohorts were retrieved from the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>) (18) and included GSE73403 (n = 69), GSE74777 (n = 107), GSE157009 (n = 249), GSE157010 (n = 235), GSE157011 (n = 484), and GSE67061 (n = 699). Based on the original cohort descriptions, GSE73403 and GSE67061 were surgical-excision LUSC cohorts, GSE74777 comprised stage I–II LUSC cases, and GSE157009, GSE157010, and GSE157011 were early-stage LUSC cohorts. Among these datasets, GSE157009, GSE157010, and GSE157011 were generated on GPL570; GSE74777 on GPL17586; and GSE73403 and GSE67061 on GPL6480. GSE73403 and GSE67061 originated from the Cancer Hospital of the Chinese Academy

of Medical Sciences (CAMS), and the clinicopathological characteristics of GSE67061 were supplemented from institutional records and are summarized in **Additional File 1**. The inclusion criteria for survival analyses were: (i) histologically confirmed LUSC, (ii) availability of transcriptomic data, and (iii) availability of prognostic information. The TCGA cohort served as the discovery cohort and was randomly divided in a 7:3 ratio into a training set (n = 332) and an internal test set (n = 142) using the `caret` package in R. The GEO cohorts were analyzed separately as independent external validation sets rather than merged into a pooled meta-cohort. TCGA RNA-seq expression data were analyzed as $\log_2(\text{FPKM} + 1)$ -transformed values, consistent with the original downloaded matrix. For the GEO cohorts, processed expression matrices supplied by GEO were used together with platform-specific annotation files for gene-level analyses. Because the discovery and validation cohorts were analyzed independently, no pooled cross-platform batch-correction procedure was applied across TCGA and GEO datasets.

Lipid metabolism-associated gene set collection

Lipid metabolism-associated genes (LAGs) were curated from the Molecular Signatures Database (MSigDB, version 7.1) (19). Four predefined lipid-related gene sets were included: Reactome phospholipid metabolism, Reactome metabolism of lipids, Hallmark fatty acid metabolism, and KEGG glycerophospholipid metabolism. In the original gene-set curation workflow, 776 genes were extracted from KEGG and Reactome pathway resources, and an additional 594 lipid metabolism-associated genes were compiled from previously published studies (20-23). After removing redundancies and intersecting the curated lipid-metabolism panel with the TCGA-LUSC transcriptome, 1,064 LAGs were retained as candidate genes for downstream analysis.

Differential-expression analysis

GEO2R was used to handle and analyze GEO data (<https://www.ncbi.nlm.nih.gov/geo/geo2r/>). Differential-expression analysis between TCGA-LUSC tumor tissues and normal lung tissues was performed using the `limma` package in R (version 4.2.1) (24). Genes meeting the thresholds of false discovery rate (FDR) < 0.05 and absolute log2 fold change ($|\log_2FC|$) > 2 were classified as lipid metabolism-associated differentially expressed genes (LADEGs). Heatmaps and volcano plots were generated using `pheatmap` and `ggplot2` for visualization of differentially expressed genes.

Protein–protein interaction network and hub-gene analysis

To explore the functional relationships among LADEGs, a protein–protein interaction (PPI) network was constructed using the STRING database (<http://string-db.org>) (25). The required interaction score was set to the highest confidence level (0.9). The resulting network was visualized in Cytoscape (26). Network modules and hub genes were identified using the MCODE framework, with the following parameters: degree cutoff ≥ 2 , node score cutoff ≥ 0.2 , K-core ≥ 2 , and maximum depth = 100.

Functional enrichment analysis of LADEGs

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were conducted using the `clusterProfiler` package (27). GO analyses included biological process (BP), cellular component (CC), and molecular function (MF) terms. Enrichment significance was determined using an FDR threshold < 0.05 . These analyses were used both to characterize LADEGs identified between tumor

and normal tissues and, later, to interpret risk-associated gene sets derived from the prognostic model.

Prognostic gene screening and signature construction

Prognostic screening was performed in the TCGA training cohort. First, univariate Cox proportional hazards regression analysis was applied to the 112 LADEGs to identify genes associated with OS. Genes passing this screening step were then subjected to least absolute shrinkage and selection operator (LASSO) Cox regression using the `glmnet` framework with 10-fold cross-validation to reduce overfitting and minimize redundant predictors (28). Candidate genes retained by LASSO were entered into a multivariate Cox regression model, from which a six-gene lipid metabolism-associated prognostic signature (LAPS) was derived. The formula for the prognostic risk score was constructed as follows:

Risk score = $\sum_{i=1}^n \text{coefficients} \times \text{expression of LAGs (i)}$

$i=1$

Risk score = expression of the gene 1 $\times \beta_1$ + expression of the gene 2 $\times \beta_2$ + ... $\beta_2\beta_n$
+ expression of the gene n $\times \beta_n$

The final risk score was calculated as a linear combination of the expression values of the six genes and their regression coefficients:

Risk score = $(0.149088008951753 \times ALOX15B \text{ expression}) + (0.00716225703232336 \times CYP24A1 \text{ expression}) + (0.0513737479543645 \times PPP2R2C \text{ expression}) + (0.0509730903699623 \times PTGIS \text{ expression}) + (-0.0334355495792042 \times SPTSSB \text{ expression}) + (0.0793998976705514 \times UGT2B17 \text{ expression})$.

Risk score calculation and patient stratification

For each patient, the LAPS risk score was calculated using the fixed coefficient formula above. Patients in the TCGA training cohort were stratified into high- and low-risk groups according to the median risk score. The same risk-score formula was then applied to the internal test cohort and each external validation cohort. In keeping with the original workflow, risk stratification was performed using a median-cutoff approach within the corresponding analysis cohort.

Internal and external validation of the prognostic model

The prognostic performance of the six-gene LAPS was evaluated in the TCGA training cohort, the TCGA internal test cohort, and the six GEO validation cohorts. OS differences between high- and low-risk groups were assessed by Kaplan–Meier survival analysis with log-rank testing. Time-dependent receiver operating characteristic (ROC) curves were generated using the `survivalROC` package, based on established time-dependent ROC methodology for censored survival data (29). The original results emphasized 5-year ROC performance in the TCGA training and internal test cohorts. An ancillary recurrence-free survival (RFS) analysis was additionally performed in GSE74777, the only validation cohort explicitly described as containing RFS information. The finalized analytical pipeline did not include calibration curves, concordance-index analysis, decision-curve analysis, or nomogram construction.

Independent prognostic analysis and clinicopathological correlation

To determine whether the LAPS risk score was independent of conventional clinicopathological variables, univariate and multivariate Cox regression analyses were performed in the TCGA cohort and, where appropriate data were available, in the external validation cohorts. Covariates included age, sex, tumor stage, and other available

clinicopathological variables in the respective datasets. In the TCGA cohort, the available variables included age, sex, overall stage, T stage, N stage, metastasis status, and selected treatment-related annotations, although some variables contained a substantial proportion of missing or unknown values.

Functional enrichment analysis of risk-associated genes

To characterize biological processes associated with the prognostic signature, TCGA-LUSC tumors were divided into high- and low-risk groups according to the median risk score, and differentially expressed genes between the two groups were identified using the thresholds of $|\log_2FC| \geq 1$ and $FDR < 0.05$. GO and KEGG enrichment analyses of these risk-associated genes were then performed using `clusterProfiler` (27). In addition, the correlation between genome-wide expression and the risk score was assessed in the TCGA cohort, and positively and negatively correlated gene sets were further explored using pathway analyses.

Immune infiltration and immune-related analyses

Tumor immune-cell composition was estimated in the TCGA-LUSC cohort using the CIBERSORT algorithm (30). Relative fractions of 22 immune-cell subsets were inferred from the TCGA gene-expression matrix, and the abundance of each immune-cell population was compared between the high- and low-risk groups. These analyses were intended as computational estimates of immune contexture rather than direct immunophenotyping.

Gene set enrichment analysis

Gene set enrichment analysis (GSEA) was performed using GSEA software (version 4.3.2) to investigate coordinated pathway-level differences between the high- and low-

risk groups (31). The TCGA-LUSC gene-expression matrix served as the input dataset. Hallmark and KEGG gene sets from MSigDB (version 7.2) were used as reference gene sets. Pathways with nominal $P < 0.05$ and FDR $q < 0.05$ were considered significantly enriched.

Immunotherapy-response-related computational analyses

Because no immunotherapy-treated LUSC cohort was available in this study, immunotherapy-related analyses were restricted to in silico exploratory evaluation. The Tumor Immune Dysfunction and Exclusion (TIDE) (<http://tide.dfci.harvard.edu>) framework was applied to normalized TCGA-LUSC RNA-seq data to derive TIDE, microsatellite instability (MSI), T-cell dysfunction, and T-cell exclusion scores (32). In addition, expression levels of selected immune checkpoint genes and immunotherapy-related biomarkers, including PDCD1, CD274, CTLA4, HAVCR2, LAG3, IFNG, Merck18, and cancer-associated fibroblast (CAF)-related scores, were compared with the LAPS risk score in exploratory association analyses. These analyses were used to examine immune-related correlates of the model and were not intended to constitute clinical validation of immune-checkpoint response.

Candidate compound prediction

To generate exploratory hypotheses regarding candidate compounds associated with the LUSC lipid-metabolism signature, the Connectivity Map (CMap) (<http://www.broadinstitute.org>) platform was queried using the LUSC-associated expression signature (33). LADEGs were submitted to the CMap resource to identify perturbagens showing positive or negative connectivity with the disease-associated transcriptional pattern. Compounds were ranked according to their connectivity profiles,

and the top connected compounds were extracted for descriptive reporting. These analyses were considered hypothesis-generating and were not experimentally validated in the present study.

Public atlas-based protein-expression assessment

As an orthogonal public-resource analysis, protein-expression patterns for genes within the six-gene signature were assessed using the Human Protein Atlas (HPA, <https://www.proteinatlas.org/>) (34). Available immunohistochemistry (IHC) data were reviewed for protein expression in LUSC and normal lung tissues. Available immunofluorescence data, including subcellular localization information, and HPA lung cancer cell-line transcript-expression data were also examined. These data were used as public atlas-based supportive evidence and were not considered equivalent to in-house functional validation.

Experimental validation of selected genes in LUSC cell lines

Cell lines and culture conditions

To provide biological support for selected genes within the six-gene lipid metabolism-associated prognostic signature, in vitro experiments were performed using the normal human bronchial epithelial cell line BEAS-2B and the LUSC cell lines NCI-H1703, NCI-H520, and SK-MES-1. BEAS-2B cells were cultured in high-glucose Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (FBS) and 1% penicillin–streptomycin. NCI-H1703 and NCI-H520 cells were cultured in RPMI-1640 medium supplemented with 10% FBS and 1% penicillin–streptomycin. SK-MES-1 cells were cultured in Eagle's minimum essential medium (EMEM) supplemented with 10% FBS and 1% penicillin–streptomycin. Cells were maintained at

37°C in a humidified incubator containing 5% CO₂. No evidence of mycoplasma contamination was noted during the study period. Selected genes within the signature were first screened at baseline mRNA and protein levels in these cell lines. *ALOX15B* and *PPP2R2C* were then prioritized for siRNA-mediated knockdown, and *ALOX15B* underwent further functional evaluation by colony-formation and RSL3-related cell-viability assays.

RNA extraction, reverse transcription, and quantitative real-time PCR

Total RNA was extracted from cultured cells using TRIzol reagent (Invitrogen, Thermo Fisher Scientific; catalog no.15596026) according to the manufacturer's protocol. RNA concentration and purity were assessed spectrophotometrically using a NanoDrop 2000 instrument. Complementary DNA (cDNA) was synthesized using TransScript® Uni All-in-One First-Strand cDNA Synthesis SuperMix for qPCR (One-Step gDNA Removal) (TransGen Biotech, Beijing, China; AU341). Quantitative real-time PCR (qPCR) was performed using PerfectStart® Visual Green qPCR SuperMix (TransGen Biotech; AQ621) on an ABI 7500 Fast Real-Time PCR System according to the manufacturer's instructions. Relative mRNA expression was calculated using the 2^{-ΔΔCt} method after normalization to ACTB.

The primer sequences were as follows:

ALOX15B forward 5'-CAGTGGAAGGCTTACAACCCA-3' and reverse 5'-

CAGTGCTCAAATGCGTGCT-3';

PPP2R2C forward 5'-CCGGGTCGTCATCTTCCAG-3' and reverse 5'-

GTTGGTGGACAGGAGTGAGT-3';

343 *CYP24A1* forward 5'-ATGGTGGCCATCACTGTCC-3' and reverse 5'-

344 CACGTCGTAGTTCCCGTAGC-3';

345 *PTGIS* forward 5'-ACTGTTGCTGCTGCTGCTACTG-3' and reverse 5'-

346 GAGGAAGATGGCATAGGCATGGAAG-3';

347 *UGT2B17* forward 5'-TGCCGATGATTGGCTTGACA-3' and reverse 5'-

348 CCCAGAAAAGGAGACCAAGCA-3'; and

349 *ACTB* forward 5'-CCTTCCTGGGCATGGAGTC-3' and reverse 5'-

350 TGATCTTCATTGTGCTGGGTGG-3'.

351 Unless otherwise indicated, qPCR experiments were performed using three independent

352 biological replicates, each measured in technical triplicate.

353 **siRNA design and transfection**

354 To investigate the functional relevance of selected signature genes, siRNA-mediated

355 knockdown experiments were performed for *ALOX15B* and *PPP2R2C*. A non-targeting

356 siRNA was used as the negative control (Si-NC).

357 The siRNA sequences were as follows:

358 *ALOX15B-Si1*, sense 5'-GCUUCUCUGAGUUGAUACATT-3' and antisense 5'-

359 UGUAUCAACUCAGAGAAGCTT-3';

360 *ALOX15B-Si2*, sense 5'-GGAACGCUUUGUCUCUGAATT-3' and antisense 5'-

361 UUCAGAGACAAAGCGUUCCTT-3';

362 *PPP2R2C-Si1*, sense 5'-GCCUGGAGAUAGAGGAGAATT-3' and antisense 5'-

363 UUCUCCUCUAUCUCCAGGCTT-3'; and

364 *PPP2R2C-Si2*, sense 5'-CCAUCACCGACAGGAGCUUTT-3' and antisense 5'-

365 AAGCUCCUGUCGGUGAUGGTT-3'.

Cells at approximately 60%–80% confluence were transfected with siRNAs at a final concentration of 25 nM using Lipofectamine RNAiMAX (Invitrogen, Thermo Fisher Scientific; catalog no. 13778075) according to the manufacturer's instructions. The transfection medium was replaced with fresh complete medium after 24 h, and knockdown efficiency was assessed by qPCR at 72 h post-transfection. Based on the experimental records, *ALOX15B* expression was reduced by approximately twofold after knockdown, whereas *PPP2R2C* expression was reduced to a nearly undetectable level by qPCR.

Protein extraction and western blotting

Whole-cell lysates were prepared from BEAS-2B, NCI-H1703, NCI-H520, and SK-MES-1 cells using RIPA lysis buffer supplemented with protease and phosphatase inhibitors. Protein concentration was determined using a BCA protein assay. Equal amounts of total protein (20–40 µg per lane) were separated by 10%–12% SDS-PAGE and transferred onto PVDF membranes. Membranes were blocked with 5% non-fat milk or bovine serum albumin in TBST for 1–2 h at room temperature and then incubated overnight at 4°C with primary antibodies against ALOX15B (Abcam, ab23691, 1:1,000), PPP2R2C (Proteintech, 12747-1-AP, 1:500–1:2,000), PTGIS (Abcam, ab23668, 1:1,000), UGT2B17 (Thermo Fisher, PA5-13430, 1:500–1:1,000), and ACTB (Cell Signaling Technology, #4970, 1:2,000). After washing, membranes were incubated with HRP-conjugated anti-rabbit IgG secondary antibody (Cell Signaling Technology, #7074, 1:3,000) and visualized using an enhanced chemiluminescence detection system on a Bio-Rad ChemiDoc MP imaging platform. ACTB served as the loading control. Western

blotting was used to support differential protein expression of selected genes across normal epithelial and LUSC cell lines.

Colony-formation assay

To assess the effect of *ALOX15B* suppression on clonogenic growth, colony-formation assays were performed in NCI-H1703 cells after siRNA-mediated knockdown. Cells were first transfected with Si-NC, Si-*ALOX15B*-1, or Si-*ALOX15B*-2, and after 24–48 h were trypsinized, counted, and reseeded at low density in 6-well plates at approximately 500–1,000 cells per well. Cells were then cultured for 10–14 days until visible colonies formed. Colonies were fixed with 4% paraformaldehyde for 15 min at room temperature, stained with crystal violet, photographed, and compared among the three groups.

RSL3-related cell-viability assay

To investigate whether *ALOX15B* suppression influenced responses to ferroptosis-related stress, RSL3-related cell-viability assays were performed in NCI-H520 and NCI-H1703 cells after siRNA-mediated *ALOX15B* knockdown. After transfection, cells were seeded into 96-well plates at approximately 5×10^3 to 1×10^4 cells per well and treated with serial concentrations of RSL3 (10^{-9} to 10^{-4} M; Selleck Chemicals, catalog no. S8155) for 24 h. Relative cell viability was then assessed using a CCK-8 assay kit (Dojindo Molecular Technologies, catalog no. CK04), with absorbance measured at 450 nm using a Thermo Scientific Multiskan GO microplate reader. Dose–response curves were plotted as relative viability versus log₁₀ RSL3 concentration. Viability values were normalized to the control condition. These experiments were interpreted as an assessment of RSL3-related sensitivity rather than definitive mechanistic proof of ferroptosis regulation.

Statistical analysis

Bioinformatic analyses were performed primarily in R (version 4.2.1) using the `limma`,
`pheatmap`, `ggplot2`, `caret`, `glmnet`, `survival`, `survivalROC`, and `clusterProfiler`
packages, together with GSEA software (version 4.3.2) for enrichment analysis.
Experimental data visualization and selected statistical analyses were performed using
GraphPad Prism (version 9.0). Data are presented as mean $\pm$ standard deviation (SD)
unless otherwise specified. Continuous variables were compared using Student's t-test or
the Wilcoxon rank-sum test, as appropriate. Differences between two groups were
analyzed using Student's t-test, whereas comparisons among multiple groups were
evaluated using one-way analysis of variance (ANOVA) as appropriate. Survival
differences were evaluated using Kaplan–Meier analysis and the log-rank test. Univariate
and multivariate Cox proportional hazards regression analyses were used to identify
independent prognostic factors. All statistical tests were two-sided, and $P < 0.05$ was
considered statistically significant unless otherwise specified. For genome-scale analyses,
multiple testing was controlled using the false discovery rate. Unless otherwise stated, the
experiments were performed using three independent biological replicates, with qPCR
reactions run in technical triplicate.

Ethical statement

This study was approved by the Institutional Review Board of the Cancer Hospital,
Chinese Academy of Medical Sciences (CAMS), which waived the requirement for
informed consent because the clinical transcriptomic analyses were based exclusively on
publicly available, de-identified datasets. No additional patient-derived tissue specimens
were collected for the in vitro validation experiments described in this study. All

procedures were conducted in accordance with relevant institutional policies and applicable guidelines.

Results

Identification of lipid metabolism-associated differentially expressed genes in LUSC

A total of 1,064 lipid metabolism-associated genes (LAGs) were identified by intersecting the curated lipid metabolism gene set with the TCGA-LUSC transcriptome. The full list of these candidate genes is provided in **Figure 2A and Additional File 2**. Differential-expression analysis of 502 tumor samples and 49 normal lung tissues in the TCGA cohort identified 112 lipid metabolism-associated differentially expressed genes (LADEGs), including 63 upregulated and 49 downregulated genes **Figures 2B, 2C**. The complete list of LADEGs is provided in **Additional File 3**. These transcriptional differences were visualized by heatmap and volcano plot and were carried forward for downstream functional and prognostic analyses. Baseline clinicopathological characteristics of the TCGA-LUSC cohort are summarized in **Table 1**.

Functional enrichment and network characteristics of candidate genes

Functional enrichment analysis of the 112 LADEGs indicated that these genes were mainly involved in lipid-related metabolic processes. GO analysis showed enrichment in fatty acid metabolic process, lipid catabolic process, icosanoid metabolic process, glycerolipid metabolic process, lipid localization, and oxidoreductase activity (**Figure 3A**). KEGG analysis further indicated enrichment in arachidonic acid metabolism, ether lipid metabolism, glycerophospholipid metabolism, metabolism of xenobiotics by

cytochrome P450, and steroid biosynthesis-related pathways (**Figure 3B**). The detailed KEGG pathway output is provided in **Additional File 4**.

To further characterize interactions among LADEGs, a high-confidence protein–protein interaction network was constructed (**Figure 4A and 4B**). Cytoscape-based network analysis identified the top 15 hub genes, including EHHADH, ASAH1, LPL, PTGES2, CBR1, TBL1X, PTGES, AKR1C3, CARM1, PTGS2, MAOA, ALOX5, TBL1XR1, ACSL1, and ME1. The interaction-network output is shown in **Figure 4C**, and the hub-gene list is provided in **Additional File 5**.

These network-level findings supported the view that the differentially expressed lipid-associated genes participate in interconnected metabolic and cancer-related processes rather than acting as isolated transcriptional events.

Construction of the six-gene lipid metabolism-associated prognostic signature

Among the 112 LADEGs, univariate Cox regression in the TCGA training cohort identified 15 genes associated with overall survival; these genes are listed in **Additional File 6**. LASSO regression further reduced the candidate set to 11 genes, which are provided in **Figure 5A, 5B and Additional File S7**. Multivariate Cox regression ultimately yielded a six-gene lipid metabolism-associated prognostic signature (LAPS) composed of *ALOX15B*, *CYP24A1*, *PPP2R2C*, *PTGIS*, *SPTSSB*, and *UGT2B17* (**Figure 5C**). The multivariable model outputs for the final six-gene signature are summarized in **Additional File 8**.

Risk score = expression of the *ALOX15B* $\times$ 0.149088008951753 + expression of the *CYP24A1* $\times$ 0.00716225703232336 + expression of the *PPP2R2C* $\times$

0.0513737479543645 +expression of the *PTGIS* $\times$ 0.0509730903699623 +expression of
the *SPTSSB* $\times$ -0.0334355495792042 +expression of the *UGT2B17* $\times$
0.0793998976705514.

Risk-score distribution, survival-status plots, and heatmaps in the TCGA training cohort
showed a progressive increase in death events with increasing risk score **Figure 6A and**
6B . The expression pattern of the signature was also directionally consistent with the
final model: *SPTSSB* was relatively lower in the high-risk group, whereas *ALOX15B*,
CYP24A1, *PPP2R2C*, *PTGIS*, and *UGT2B17* tended to be higher **Figure 6C**. These
observations supported the internal coherence of the six-gene model before formal
validation analyses.

Prognostic stratification performance in the TCGA training cohort

Kaplan–Meier analysis in the TCGA training cohort showed that patients classified as
high risk had significantly poorer overall survival than those classified as low risk
(**Figure 6D**). Time-dependent ROC analysis demonstrated moderate prognostic
discrimination, with a 5-year area under the curve (AUC) of 0.647 in the training cohort
(**Figure 6E**). These findings indicate that the six-gene LAPS was able to stratify survival
risk in the discovery cohort with moderate predictive performance.

Internal and external validation of the prognostic signature

The same six-gene formula was next evaluated in the TCGA internal test cohort (**Figure**
7). As in the training cohort, the internal test set showed increasing death events with
increasing risk score, and high-risk patients had significantly shorter overall survival than
low-risk patients (**Figure 7A-7D**). The 5-year AUC in the internal test cohort was 0.677,

indicating comparable, moderate predictive performance in the internal validation setting
(**Figure 7E**).

External validation was then performed across six GEO cohorts: GSE73403, GSE74777, GSE157009, GSE157010, GSE157011, and GSE67061. Across these cohorts, the six-gene LAPS showed a consistent trend toward poorer overall survival in the high-risk group, and the expression pattern of the six-gene panel remained directionally consistent with that observed in the TCGA cohorts. The Kaplan–Meier survival curves for the six external cohorts are provided in **Additional File 9**. In addition, analysis of GSE74777 showed that high-risk patients also had poorer 5-year recurrence-free survival, as shown in **Additional File 10**. Collectively, these findings support the generalizability of the signature across multiple retrospective datasets and platforms.

Independent prognostic significance and clinicopathological associations

Univariate Cox regression analysis in the TCGA cohort showed that the risk score was significantly associated with overall survival ($p = 4.67\text{E-}06$) (**Figure 8A**). In multivariate Cox analysis, the risk score remained significant after adjustment for clinicopathological variables, indicating that the six-gene LAPS functioned as an independent prognostic factor in the discovery cohort ($p = <0.001$) (**Figure 8B**). The univariate Cox results for the TCGA cohort are summarized in **Additional File 11**, and the corresponding validation-cohort Cox outputs are provided in **Additional File 12**. Additional multivariable verification of the model and the clinicopathological distribution of low- and high-risk groups across validation cohorts are shown in **Supplementary Figure S13**. Clinical annotation heatmaps also showed that low- and high-risk groups were distributed differently with respect to tumor stage, TNM stage, age, and sex in the TCGA cohort

($P < 0.05$) (**Figure 8C**), and the external validation figures indicated cohort-specific relationships between risk score and clinicopathological factors. These findings support an association between the six-gene LAPS and established disease characteristics, while also indicating that the risk score retains prognostic value beyond conventional clinicopathological grouping alone.

Biological pathway characteristics associated with the risk groups

To further characterize the biology associated with the prognostic model, genes correlated with the risk score were analyzed in the TCGA cohort. This analysis identified **707 genes negatively correlated** and **638 genes positively correlated** with the LAPS risk score (**Figure 9A, 9B and Additional File 14**). GO and KEGG analyses of these risk-associated genes showed enrichment in immune- and metabolism-related pathways, including B-cell-mediated immunity, positive regulation of lymphocyte activation, phagocytosis, humoral immune response, complement activation, oxidative phosphorylation, and reactive oxygen species-related carcinogenic pathways (**Figure 9C, 9D and Additional File 15**).

Immune infiltration and immune-related differences between risk groups

Because pathway analyses indicated substantial immune-related differences between risk groups, immune-cell composition was further estimated using CIBERSORT in the TCGA cohort. Six immune-cell populations differed between the high- and low-risk groups. Compared with the low-risk group, the high-risk group showed lower levels of CD8⁺ T cells, follicular helper T cells, and eosinophils, but higher levels of resting memory CD4⁺ T cells, M0 macrophages, and neutrophils (**Figure 10 and Additional File 16**). These

findings suggest that the six-gene signature is associated with distinct immune-microenvironment characteristics in LUSC.

Gene Set Enrichment Analysis

Gene set enrichment analysis (GSEA) further demonstrated distinct pathway patterns between the two risk groups (**Additional File 17**). The high-risk group, significantly correlated with KEGG pathways, was enriched in PPAR signaling, mTOR signaling, WNT signaling, regulation of autophagy, VEGF signaling, RIG I receptor signaling pathway, prostate cancer, fatty acid metabolism, and drug metabolism by cytochrome P450 (**Figure 11A**). In contrast, the low-risk group was enriched in nod-like receptor signaling, mismatch repair, chemokine signaling, JAK–STAT signaling, cytokine–cytokine receptor interaction, TGF- β signaling, p53 signaling, and apoptosis (**Figure 11B**). GO enrichment analysis revealed that the high-risk group was related to an increase in the activity of various biological processes and molecular functions, such as negative regulation of autophagosome assembly, the glucocorticoid metabolic process, the cellular carbohydrate catabolic process, the regulation of cell growth involved in cardiac muscle cell development, aspartic type peptidase activity, corticosteroid hormone secretion, neurotransmitter biosynthetic process, and negative regulation of smooth muscle cell differentiation (**Figure 11C**). Furthermore, regulation of T-cell migration, regulation of the activin receptor signaling pathway, maintenance of protein location in the nucleus, negative regulation of nucleocytoplasmic transport, DNA ligation, integrin activation, intracellular calcium-activated chloride channel activity, and protein peptidyl-prolyl isomerization, were associated with the low-risk group (**Figure 11D**). These pathway-

level differences are consistent with distinct biological states associated with the six-gene risk model.

Immunotherapy-response-related exploratory analyses

Exploratory in silico analyses were performed to examine whether the risk groups differed in immunotherapy-response-related biomarkers (**Additional File 18**). In the TCGA cohort, the high-risk group showed higher IFNG, T-cell dysfunction, Merck18, and CAF scores, and TIM-3 (HAVCR2) expression was positively correlated with the risk score and elevated in the high-risk group (**Figure 12 A–E**). In contrast, LAG3 expression was higher in the low-risk group (**Figure 12F**). These findings indicate that the two risk groups differed in selected immune-checkpoint- and dysfunction-related features.

However, several central associations were not significant. No significant relationship was identified between the risk score and overall TIDE score, CD274 score, or T-cell exclusion score, and the risk score was not associated with PDCD1, CD274, or CTLA4 expression (**Additional File 19**). Accordingly, the immunotherapy-related analyses are best interpreted as exploratory in silico associations rather than direct evidence of clinical response prediction.

Public protein-expression support

As an orthogonal public-resource analysis, protein-expression patterns for genes within the six-gene signature were examined using the Human Protein Atlas (HPA). In the available HPA immunohistochemistry data, SPTSSB was not represented. ALOX15B, PPP2R2C, and PTGIS showed higher protein expression in LUSC than in normal lung tissue, whereas no clear statistical support was described for CYP24A1 and UGT2B17 in

the original dataset summary (**Figure 13 and Additional File 20**). The HPA resource also provided an immunofluorescence image indicating mitochondrial localization of CYP24A1 in A549 cells (**Additional File 21**) and transcript-expression information for the six genes across lung cancer cell lines (**Figure 14 and Additional File 22**). These analyses provided public protein-level context for the signature but were not equivalent to in-house experimental validation.

Candidate compound prediction

To generate treatment-related hypotheses, the LADEG signature was queried against the Connectivity Map (CMap) database (<http://www.broadinstitute.org>). This analysis yielded the top connected compounds associated with the LUSC lipid metabolism-associated expression pattern (**Figure 15 and Additional File 23**). In the original CMap framework, strongly negative connectivity scores (*tau*) indicate compounds that may reverse the disease-associated transcriptional pattern and therefore warrant further investigation. In the present study, these compounds are best regarded as putative candidates for future experimental testing rather than validated therapeutic agents.

Experimental validation of selected genes within the six-gene signature

To strengthen the biological support for the six-gene lipid metabolism-associated prognostic signature, selected genes within the model were experimentally examined in LUSC cell lines (**Figure 16A–J**). Baseline qPCR analysis demonstrated that the expression levels of selected signature genes differed across BEAS-2B, NCI-H1703, NCI-H520, and SK-MES-1 cells. *ALOX15B* showed relatively higher expression in NCI-H1703 and NCI-H520 than in BEAS-2B, whereas *CYP24A1* and *UGT2B17* were most strongly expressed in NCI-H520. *PPP2R2C* and *PTGIS* also displayed heterogeneous

expression patterns across the tested LUSC cell lines, with increased expression in tumor-derived cells relative to the normal epithelial control in selected contexts (**Figure 16B, D, F, G, and I**). Western blotting further supported differential protein expression of ALOX15B, PPP2R2C, PTGIS, and UGT2B17 across BEAS-2B, NCI-H1703, NCI-H520, and SK-MES-1 cells (**Figure 16J**).

We next performed siRNA-mediated knockdown of *ALOX15B* and *PPP2R2C* to validate selected components of the signature experimentally. qPCR confirmed effective suppression of both target genes after siRNA treatment. *ALOX15B* expression was markedly reduced after transfection with two independent siRNAs (**Figure 16H**), whereas *PPP2R2C* expression decreased to a nearly undetectable level by qPCR in parallel knockdown experiments. Functionally, *ALOX15B* knockdown reduced colony-forming capacity in NCI-H1703 cells compared with the negative-control group (**Figure 16A**). In addition, RSL3-related viability assays in both NCI-H520 and NCI-H1703 cells showed altered viability responses after *ALOX15B* knockdown across the tested dose range (**Figure 16C and E**). Together, these findings provide biological support for selected genes within the bioinformatically derived signature and strengthen the biological plausibility of the lipid metabolism-associated prognostic framework, with the strongest functional evidence centered on *ALOX15B*.

Discussion

The present study identifies a lipid metabolism-associated prognostic framework for LUSC that is supported by both multi-cohort retrospective transcriptomic analyses and selected in vitro validation. In a disease setting where clinically meaningful molecular

stratification remains limited, a biologically grounded model that links survival differences with metabolism-associated transcriptional programs is of clear translational interest (35). In this context, the six-gene signature described here provides a focused risk-stratification approach centered on lipid metabolism and is strengthened by experimental support for selected components of the model, particularly *ALOX15B* and *PPP2R2C*.

Several aspects of the main prognostic finding warrant emphasis. First, the six-gene lipid metabolism-associated prognostic signature (LAPS) consistently separated patients with different survival outcomes in the TCGA training cohort, the TCGA internal test cohort, and six external GEO cohorts. Second, the risk score remained significant in multivariable analyses, supporting its value beyond conventional clinicopathological variables in the available datasets. Third, the model showed moderate rather than exceptional discrimination, with 5-year AUCs of 0.647 and 0.677 in the training and internal test cohorts, respectively. These values support prognostic stratification, but they do not justify presenting the model as a clinically ready decision tool. This more restrained interpretation is especially important in light of current prediction-model standards, which increasingly emphasize careful distinction between model development, retrospective evaluation, and true clinical applicability.

This balanced positioning is also necessary because the field is already populated by metabolism-oriented LUSC signatures. Prior studies have reported metabolism-related and fatty acid metabolism-related prognostic models in LUSC with survival stratification, immune analyses, and multi-cohort validation (16, 36). Against that background, the value of the present study does not lie in simply introducing another transcriptome-

derived signature. Rather, its contribution is the combination of a lipid metabolism-focused framework with broader retrospective evaluation and selected experimental support, thereby moving the work beyond a purely public-database association study. In other words, the relevance of the manuscript is not that it resolves LUSC biomarker development, but that it attempts to connect risk stratification with biologically meaningful metabolic context and focused laboratory evidence.

The biological plausibility of the model is reinforced by the pathway-level findings. The six-gene signature emerged from genes enriched in arachidonic acid metabolism, ether lipid metabolism, glycerophospholipid metabolism, and related oxidative and steroid-associated pathways, while the risk groups differed in PPAR signaling, mTOR signaling, WNT signaling, drug metabolism by cytochrome P450, oxidative phosphorylation, and several immune-associated programs. These observations are consistent with a broader literature showing that lipid metabolic reprogramming supports membrane synthesis, redox control, stress adaptation, and reciprocal interactions with the tumor microenvironment (7). In thoracic malignancies, this biology is especially relevant because metabolic remodeling may influence both tumor-cell fitness and the broader ecological state of the tumor microenvironment, even when the exact causal direction of a given transcriptomic signature remains unresolved.

At the level of individual genes, the present study provides differing depths of evidence. *ALOX15B* deserves particular attention because it entered the model computationally, showed differential expression in LUSC cell lines, underwent successful siRNA-mediated suppression, and was the only gene for which colony-formation and RSL3-related viability assays were performed (37). As a lipoxygenase family member involved

in lipid oxidation, *ALOX15B* is biologically relevant to redox-active lipid metabolism, and recent work in other tumor systems has linked *ALOX15B* activity to ferroptosis-related biology (38-40). Our data do not establish a ferroptosis mechanism in LUSC, but they do make a ferroptosis-related interpretation biologically plausible and worthy of deeper mechanistic investigation.

PPP2R2C should be interpreted more cautiously. In this study, *PPP2R2C* contributed to the prognostic model, showed differential baseline expression, and underwent effective siRNA-mediated knockdown by qPCR, but it did not receive the same degree of functional follow-up as *ALOX15B*. Biologically, *PPP2R2C* encodes a regulatory subunit of PP2A, a phosphatase complex with broad relevance to oncogenic signaling control and cancer cell-state regulation (41, 42). This background supports the plausibility of *PPP2R2C* as a biologically meaningful component of the signature, but the present study does not define its mechanistic role in LUSC. The same caution applies to the remaining genes within the panel: their inclusion in the model is supported by statistical and, for selected genes, experimental evidence, but not all were validated to an equivalent biological depth.

The immune-related analyses add an informative but necessarily restrained dimension to the study. High- and low-risk groups differed in estimated immune-cell composition, including CD8+ T cells, follicular helper T cells, M0 macrophages, neutrophils, and eosinophils, and they also differed in selected immune-checkpoint- and dysfunction-related features such as IFNG, TIM-3, Merck18, CAF, and LAG3. These findings suggest that the lipid metabolism-associated signature may reflect differences in immune contexture rather than survival alone. Such an interpretation is consistent with increasing

evidence that metabolic programs and immune-cell states are tightly intertwined in tumors (7).

However, the immunotherapy-related findings should not be overstated. A key strength of the revised manuscript is that it now distinguishes immune-feature association from genuine response prediction. In particular, the supplementary analyses indicate that the risk score was not significantly associated with overall TIDE score, CD274 score, T-cell exclusion, PDCD1, CD274, or CTLA4. These negative findings materially limit any claim that the model predicts immune-checkpoint inhibitor response in patients. More broadly, recent reviews of computational immunogenomic biomarkers emphasize that transcriptome-based predictors can reveal informative associations, but their reproducibility and clinical applicability remain limited without validation in appropriately treated cohorts (43, 44). Accordingly, our data are best interpreted as showing that the six-gene LAPS is associated with immune-related features relevant to the LUSC microenvironment, while leaving its value for actual immunotherapy prediction unresolved.

The candidate-compound analysis should also be viewed through a hypothesis-generating lens. Connectivity-based approaches such as CMap can be useful for nominating compounds whose transcriptional effects may oppose a disease-associated expression state, thereby generating testable therapeutic hypotheses. In the current study, the top connected compounds provide a starting point for future functional work, but they do not constitute validated therapeutic options for LUSC.

An important strength of the present study is that the computational findings were not left entirely unsupported biologically. In addition to the retrospective transcriptomic analyses,

we experimentally validated selected genes within the six-gene signature in LUSC cell models. Baseline qPCR and western blotting supported differential expression of several signature genes across normal bronchial epithelial and LUSC cell lines. Moreover, siRNA-mediated knockdown confirmed effective suppression of *ALOX15B* and *PPP2R2C*, and the subsequent functional assays indicated that *ALOX15B* knockdown reduced colony-forming capacity and altered RSL3-related viability responses in LUSC cells. These data do not constitute full functional validation of the complete six-gene prognostic model, and not all genes in the signature underwent equivalent experimental assessment; however, they provide important biological support for selected components of the model and strengthen the overall plausibility of a lipid metabolism-associated prognostic framework in LUSC.

Another strength of the study is its multi-cohort design. The use of TCGA for discovery, an internal TCGA split for secondary confirmation, and six independent GEO cohorts for external validation reduces the risk that the reported associations are entirely cohort-specific. The manuscript also benefits from its focus on a biologically coherent axis — lipid metabolism — rather than an agnostic genome-wide signature alone. Although this does not eliminate the limitations inherent to retrospective transcriptomic modeling, it does improve interpretability and enhances the translational appeal of the work.

Limitations

Several limitations should be acknowledged. First, although we performed experimental validation for selected genes within the six-gene signature, the full prognostic model was not comprehensively validated functionally, and not all six genes underwent equivalent experimental assessment. Second, the current functional data were more extensive for

ALOX15B than for the other genes, and further mechanistic work is required to clarify how individual signature genes contribute to LUSC progression, lipid metabolism, and ferroptosis-related biology. Third, the prognostic model was derived and evaluated retrospectively using public datasets, and its clinical utility requires prospective validation. Fourth, the immune and immunotherapy-related analyses remain indirect and exploratory, because no independent immunotherapy-treated LUSC cohort was available for direct response validation. Fifth, the candidate-compound analysis was hypothesis-generating only and was not followed by pharmacologic testing. Finally, while the RSL3-related assays suggest possible ferroptosis-relevant biology, they do not establish a full ferroptosis mechanism in LUSC.

Future Directions / Future Perspectives

Several next steps could substantially strengthen the translational trajectory of this work. Prospective validation in clinically annotated LUSC cohorts will be important to determine whether the six-gene LAPS retains prognostic value under real-world sampling, treatment, and follow-up conditions. Parallel evaluation in immunotherapy-treated populations would be particularly valuable, given that the present immune-related analyses are suggestive but not predictive. Mechanistically, deeper investigation of *ALOX15B* and *PPP2R2C* should be prioritized, including rescue experiments, lipid peroxidation readouts, ferroptosis-specific assays, and broader pathway interrogation to clarify whether these genes act as passive biomarkers, functional contributors, or both. More broadly, integrating transcriptomic signatures with proteomic, spatial, or single-cell data could help determine whether the current risk groups correspond to reproducible metabolic-immune states in LUSC. Finally, the candidate compounds identified by CMap

merit stepwise experimental testing, but ideally within models informed by the biology of the signature rather than by transcriptional reversal alone (43, 44). In summary, this study supports the existence of a lipid metabolism-associated prognostic framework in LUSC that is reproducible across multiple retrospective cohorts and biologically reinforced by selected experimental validation. The six-gene signature should be interpreted as a prognostic and biology-linked model rather than a clinically deployed therapeutic tool, but it nevertheless provides a useful platform for future work at the interface of cancer metabolism, tumor microenvironment biology, and translational biomarker development. By coupling multi-cohort human data with focused wet-lab support, the study advances a more credible and biologically interpretable direction for lipid metabolism-associated biomarker research in LUSC.

Conclusion

In conclusion, this study presents a lipid metabolism-associated six-gene prognostic framework for lung squamous cell carcinoma that is supported by multi-cohort retrospective transcriptomic analysis and experimental evaluation of selected genes within the signature. The model was associated with survival stratification and distinct metabolic and immune-related features, while the in vitro findings provided biological support for selected components of the framework, particularly *ALOX15B* and *PPP2R2C*. These results offer a biologically informed basis for future translational investigation and warrant further prospective validation and mechanistic study in clinically annotated LUSC cohorts.

797 **List of Abbreviations**

- 798 **ACTB:** Actin beta
- 799 **AUC:** Area under the curve
- 800 **BP:** Biological process
- 801 **CAF:** Cancer-associated fibroblast
- 802 **CAMS:** Cancer Hospital of the Chinese Academy of Medical Sciences
- 803 **CC:** Cellular component
- 804 **CIBERSORT:** Cell-type Identification by Estimating Relative Subsets of RNA
- 805 Transcripts
- 806 **CMap:** Connectivity Map
- 807 **DEGs:** Differentially expressed genes
- 808 **FDR:** False discovery rate
- 809 **FC:** Fold change
- 810 **FPKM:** Fragments per kilobase of transcript per million mapped reads
- 811 **GEO:** Gene Expression Omnibus
- 812 **GO:** Gene Ontology
- 813 **GSEA:** Gene set enrichment analysis
- 814 **HPA:** Human Protein Atlas
- 815 **ICI:** Immune checkpoint inhibitor
- 816 **IHC:** Immunohistochemistry
- 817 **KEGG:** Kyoto Encyclopedia of Genes and Genomes
- 818 **LADEGs:** Lipid metabolism-associated differentially expressed genes
- 819 **LAGs:** Lipid metabolism-associated genes

820 **LAPS:** Lipid metabolism-associated prognostic signature
821 **LUAD:** Lung adenocarcinoma
822 **LUSC:** Lung squamous cell carcinoma
823 **MF:** Molecular function
824 **MSI:** Microsatellite instability
825 **MSigDB:** Molecular Signatures Database
826 **mTOR:** Mechanistic target of rapamycin
827 **NSCLC:** Non-small cell lung cancer
828 **OS:** Overall survival
829 **PPAR:** Peroxisome proliferator-activated receptor
830 **PPI:** Protein–protein interaction
831 **qPCR:** Quantitative real-time polymerase chain reaction
832 **RFS:** Recurrence-free survival
833 **ROC:** Receiver operating characteristic
834 **siRNA:** Small interfering RNA
835 **STRING:** Search Tool for the Retrieval of Interacting Genes/Proteins
836 **TCGA:** The Cancer Genome Atlas
837 **TIDE:** Tumor Immune Dysfunction and Exclusion
838 **VEGF:** Vascular endothelial growth factor

839

840 **Declarations**

841 *Ethics approval and consent to participate*

This study was approved by the Institutional Review Board of the Cancer Hospital, Chinese Academy of Medical Sciences (CAMS). The requirement for informed consent was waived because the clinical transcriptomic analyses were based exclusively on publicly available, de-identified datasets. No additional patient-derived tissue specimens were collected for the in vitro validation experiments.

Consent for publication

Not applicable

Availability of data and materials

The datasets and public resources used in this study are available from TCGA, GEO, MSigDB, STRING, GSEA, TIDE, HPA, and CMap, and the supporting data are available within the article and its supplementary materials.

Competing interests

The authors declare that they have no competing interests

Funding

Not applicable

Authors' contributions

SM, XJM, MMD, MB, YG, and HJ conceptualized and designed the study and reviewed and revised the manuscript. SM, XJM, MMD, MB and HJ were involved in data acquisition and analysis. SM, XJM, MMD, MB and HJ supervised the study. All authors read and approved the final manuscript.

Acknowledgments

Not applicable

Additional Materials

Supplementary materials for this article include additional figures, tables, file legends, and supporting datasets related to the bioinformatic analyses, external validation, and experimental procedures.

Reference List

1. Siegel RL, Kratzer TB, Wagle NS, Sung H, Jemal A. Cancer statistics, 2026. *Ca.* 2026;76(1):e70043.
2. Chen Q, Zheng X, Li J. Landscape of targeted therapies for lung squamous cell carcinoma. *Frontiers in Oncology.* 2024;14:1467898.
3. Laguna JC, Tagliamento M, Lambertini M, Hiznay J, Mezquita L. Tackling non-small cell lung cancer in young adults: from risk factors and genetic susceptibility to lung cancer profile and outcomes. *American Society of Clinical Oncology Educational Book.* 2024;44(3):e432488.
4. Laguna JC, Garcia-Pardo M, Alessi J, Barrios C, Singh N, Al-Shamsi HO, et al. Geographic differences in lung cancer: focus on carcinogens, genetic predisposition, and molecular epidemiology. *Therapeutic advances in medical oncology.* 2024;16:17588359241231260.
5. Megyesfalvi Z, Gay CM, Popper H, Pirker R, Ostoros G, Heeke S, et al. Clinical insights into small cell lung cancer: Tumor heterogeneity, diagnosis, therapy, and future directions. *CA: a cancer journal for clinicians.* 2023;73(6):620-52.

- 886 6. Hanahan D. Hallmarks of cancer: new dimensions. *Cancer discovery*.
887 2022;12(1):31-46.
- 888 7. Jin H-R, Wang J, Wang Z-J, Xi M-J, Xia B-H, Deng K, et al. Lipid metabolic
889 reprogramming in tumor microenvironment: from mechanisms to therapeutics.
890 *Journal of hematology & oncology*. 2023;16(1):103.
- 891 8. Liu R, Wang C, Tao Z, Hu G. Lipid metabolism reprogramming in cancer:
892 insights into tumor cells and immune cells within the tumor microenvironment.
893 *Biomedicines*. 2025;13(8):1895.
- 894 9. Pascual G, Benitah SA. Lipids in the tumor microenvironment: immune
895 modulation and metastasis. *Frontiers in Oncology*. 2024;14:1435480.
- 896 10. He X, Huang T, Wang Z, Bu F, Cao X, Xu M, et al. Stress-responsive membrane
897 proteins as execution nodes of tumor cell adaptation to microenvironmental stress.
898 *Oncogene*. 2026:1-15.
- 899 11. Wang J, He Y, Hu F, Hu C, Sun Y, Yang K, et al. Metabolic reprogramming of
900 immune cells in the tumor microenvironment. *International Journal of Molecular*
901 *Sciences*. 2024;25(22):12223.
- 902 12. Schneider JL, Han S, Nabel CS. Fuel for thought: targeting metabolism in lung
903 cancer. *Translational Lung Cancer Research*. 2024;13(12):3692-717.
- 904 13. Wang C, Lei M, Wang W, Jiang Y, Zhang J, Zhang Y, et al. The rising influence of
905 lipid metabolism in lung cancer: a global research perspective. *Frontiers in*
906 *oncology*. 2025;15:1562621.

907 14. Yang C, Cai H, Gao P, Xu F, Wang Q. Energy Metabolism in Non-Small Cell
908 Lung Cancer: From Hallmarks to Therapeutic Opportunities. Available at SSRN
909 4612633.

910 15. Lei K, Liang R, Tan B, Li L, Lyu Y, Wang K, et al. Effects of lipid metabolism-
911 related genes *PTGIS* and *HRASLS* on phenotype, prognosis, and tumor immunity in
912 lung squamous cell carcinoma. *Oxidative medicine and cellular longevity*.
913 2023;2023(1):6811625.

914 16. Xu J, Yu M, Fan W, Feng J. A novel gene signature related to fatty acid
915 metabolism predicts prognosis, immune landscape, and drug sensitivity in early-
916 stage lung squamous cell carcinoma. *Translational Cancer Research*.
917 2024;13(2):525-41.

918 17. Wu R, Ma R, Duan X, Zhang J, Li K, Yu L, et al. Identification of specific
919 prognostic markers for lung squamous cell carcinoma based on tumor progression,
920 immune infiltration, and stem index. *Frontiers in Immunology*. 2023;14:1236444.

921 18. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene
922 expression and hybridization array data repository. *Nucleic acids research*.
923 2002;30(1):207-10.

924 19. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The
925 molecular signatures database hallmark gene set collection. *Cell systems*.
926 2015;1(6):417-25.

- 927 20. Hu B, Yang XB, Sang XT. Construction of a lipid metabolism-related and
928 immune-associated prognostic signature for hepatocellular carcinoma. *Cancer*
929 *medicine*. 2020;9(20):7646-62.
- 930 21. Gu X, Jiang C, Zhao J, Qiao Q, Wu M, Cai B. Identification of lipid metabolism-
931 associated genes as prognostic biomarkers based on the immune
932 microenvironment in hepatocellular carcinoma. *Frontiers in Cell and*
933 *Developmental Biology*. 2022;10:883059.
- 934 22. Wang Y, Xu J, Fang Y, Gu J, Zhao F, Tang Y, et al. Comprehensive analysis of a
935 novel signature incorporating lipid metabolism and immune-related genes for
936 assessing prognosis and immune landscape in lung adenocarcinoma. *Frontiers in*
937 *Immunology*. 2022;13:950001.
- 938 23. Zhu K, Xiaoqiang L, Deng W, Wang G, Fu B. Development and validation of a
939 novel lipid metabolism-related gene prognostic signature and candidate drugs for
940 patients with bladder cancer. *Lipids in health and disease*. 2021;20(1):146.
- 941 24. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers
942 differential expression analyses for RNA-sequencing and microarray studies.
943 *Nucleic acids research*. 2015;43(7):e47-e.
- 944 25. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al.
945 STRING v11: protein–protein association networks with increased coverage,
946 supporting functional discovery in genome-wide experimental datasets. *Nucleic*
947 *acids research*. 2019;47(D1):D607-D13.

948 26. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al.
949 Cytoscape: a software environment for integrated models of biomolecular
950 interaction networks. *Genome research*. 2003;13(11):2498-504.

951 27. Xu S, Hu E, Cai Y, Xie Z, Luo X, Zhan L, et al. Using clusterProfiler to
952 characterize multiomics data. *Nature protocols*. 2024;19(11):3292-320.

953 28. Simon N, Friedman JH, Hastie T, Tibshirani R. Regularization paths for Cox's
954 proportional hazards model via coordinate descent. *Journal of statistical software*.
955 2011;39:1-13.

956 29. Heagerty PJ, Lumley T, Pepe MS. Time-dependent ROC curves for censored
957 survival data and a diagnostic marker. *Biometrics*. 2000;56(2):337-44.

958 30. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust
959 enumeration of cell subsets from tissue expression profiles. *Nature methods*.
960 2015;12(5):453-7.

961 31. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et
962 al. Gene set enrichment analysis: a knowledge-based approach for interpreting
963 genome-wide expression profiles. *Proceedings of the national academy of sciences*.
964 2005;102(43):15545-50.

965 32. Jiang P, Gu S, Pan D, Fu J, Sahu A, Hu X, et al. Signatures of T cell dysfunction
966 and exclusion predict cancer immunotherapy response. *Nature medicine*.
967 2018;24(10):1550-8.

968 33. Subramanian A, Narayan R, Corsello SM, Peck DD, Natoli TE, Lu X, et al. A
969 next generation connectivity map: L1000 platform and the first 1,000,000 profiles.
970 Cell. 2017;171(6):1437-52. e17.

971 34. Digre A, Lindskog C. The Human Protein Atlas—Spatial localization of the
972 human proteome in health and disease. Protein science. 2021;30(1):218-33.

973 35. Shen Y, Chen J-Q, Li X-P. Differences between lung adenocarcinoma and lung
974 squamous cell carcinoma: Driver genes, therapeutic targets, and clinical efficacy.
975 Genes & diseases. 2025;12(3):101374.

976 36. Ma Z, Wang L, Huang X, Ji H, Wang H, Yang Y, et al. Construction of the
977 metabolism-related models for predicting prognosis and infiltrating immune
978 phenotype in lung squamous cell carcinoma. Journal of Cancer. 2023;14(18):3539.

979 37. Liu X, Tang Y, Liu H, Zhang S, Sui h, Sun w, et al. ALOX15, a new potential
980 functional target of lung adenocarcinoma. bioRxiv. 2023:2023.03. 03.531049.

981 38. Li X, Xiong W, Wang Y, Li Y, Cheng X, Liu W. p53 activates the lipoyxygenase
982 activity of *ALOX15B* via inhibiting SLC7A11 to induce ferroptosis in bladder cancer
983 cells. Laboratory Investigation. 2023;103(5):100058.

984 39. Chen J, Zheng C, Zhang F, Sun S, Liu G, Huang G. *ALOX15B*-based
985 efferocytosis clusters: prognostic implications and immune cell infiltration in breast
986 cancer. Discover Oncology. 2025;16(1):1159.

987 40. Zhang W, Huang F, Ding X, Qin J, Wang W, Luo L. Identifying ALOX15-initiated
988 lipid peroxidation increases susceptibility to ferroptosis in asthma epithelial cells.

989 Biochimica Et Biophysica Acta (BBA)-Molecular Basis of Disease.
990 2024;1870(5):167176.

991 41. Johnson H, Narayan S, Sharma AK. Altering phosphorylation in cancer
992 through PP2A modifiers. Cancer Cell International. 2024;24(1):11.

993 42. Farrington CC. Targeted Degradation of the MYC Oncogene Using PP2A-
994 B56alpha Selective Small Molecule Modulators of Proteinphosphatase 2A as a
995 Therapeutic Strategy for Treating Myc-Driven Cancers: Case Western Reserve
996 University; 2020.

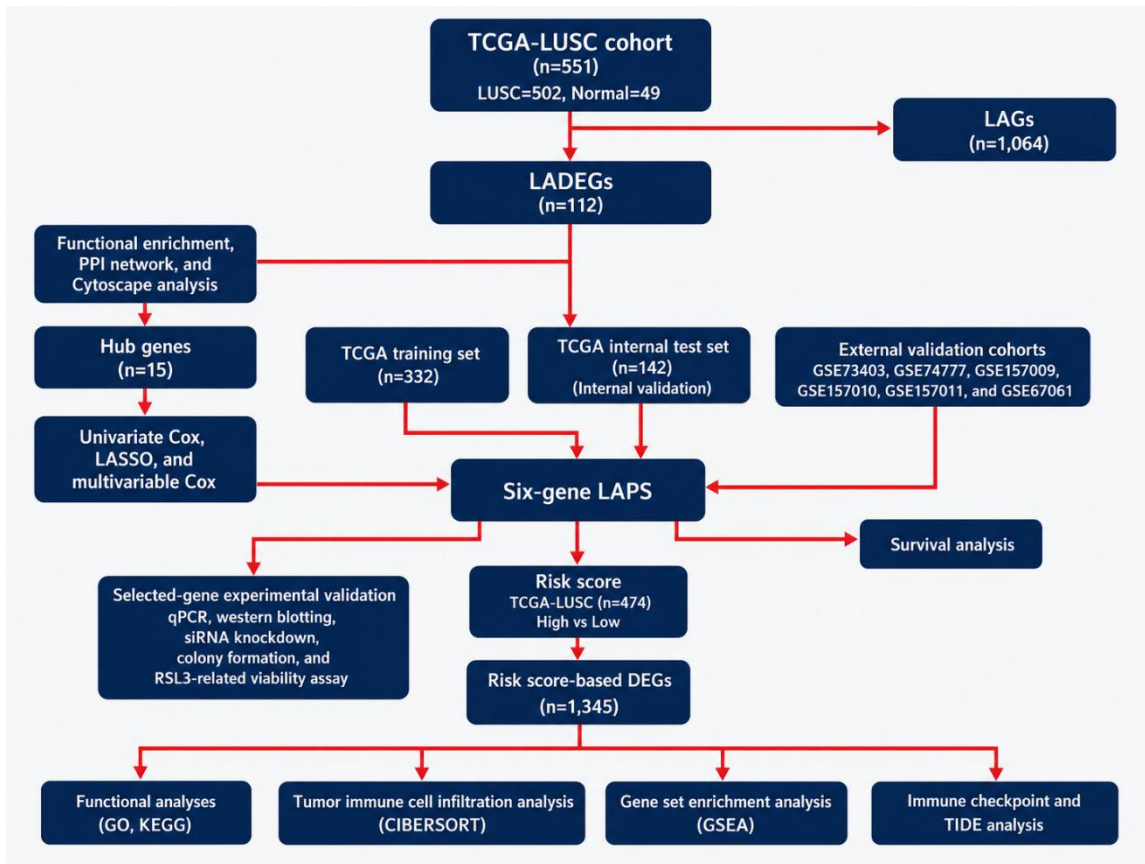
997 43. Addala V, Newell F, Pearson JV, Redwood A, Robinson BW, Creaney J, et al.
998 Computational immunogenomic approaches to predict response to cancer
999 immunotherapies. Nature Reviews Clinical Oncology. 2024;21(1):28-46.

1000 44. Wang X, Nguyen J, Nader K, Miihkinen M, Wall P, Singhal A, et al. Discovery of
1001 predictive biomarkers for cancer therapy through computational approaches.
1002 Nature Reviews Clinical Oncology. 2026:1-21.

1003 **Figures and tables and additional files**

1004 **Figures**

1005 **Figure 1. Study workflow for development, validation, and biological support of the**
1006 **lipid metabolism-associated prognostic signature in LUSC.**

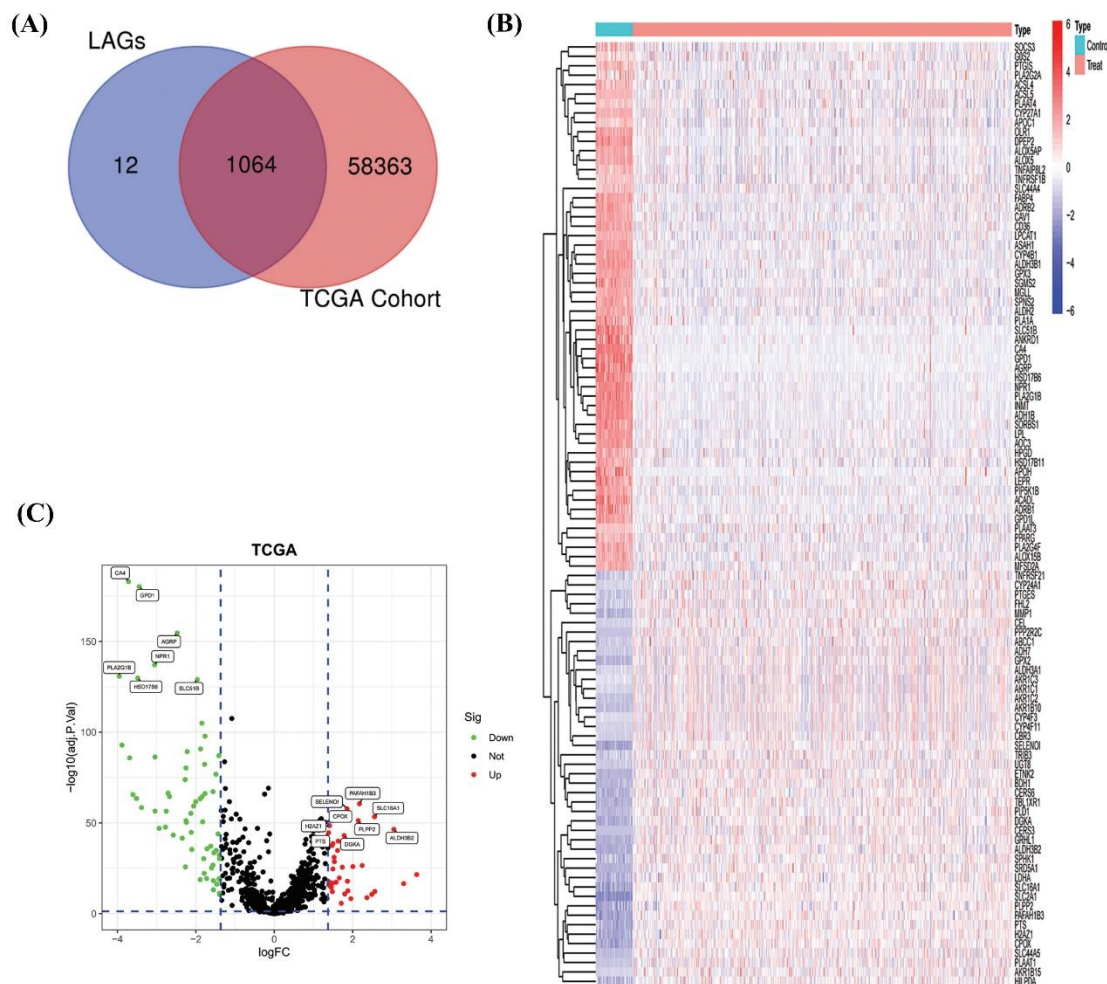


1007

1008 The overall study design is shown. Lipid metabolism-associated genes (LAGs) were
 1009 screened in the TCGA-LUSC cohort and intersected with differentially expressed genes
 1010 to identify lipid metabolism-associated differentially expressed genes (LADEGs).
 1011 Functional enrichment, protein–protein interaction (PPI) network analysis, Cytoscape-
 1012 based hub-gene identification, and Cox/LASSO-based modeling were then performed to
 1013 construct the six-gene lipid metabolism-associated prognostic signature (LAPS). The
 1014 model was evaluated in the TCGA training set, TCGA internal test set, and external GEO
 1015 validation cohorts. Downstream analyses included survival analysis, risk score-based
 1016 differential expression analysis, functional enrichment, immune-cell infiltration analysis,
 1017 gene set enrichment analysis (GSEA), and immune checkpoint/TIDE analysis. Selected
 1018 genes within the signature were further examined experimentally by qPCR, western

blotting, siRNA-mediated knockdown, colony-formation assay, and RSL3-related viability assay.

Figure 2. Identification of lipid metabolism-associated differentially expressed genes in TCGA-LUSC.

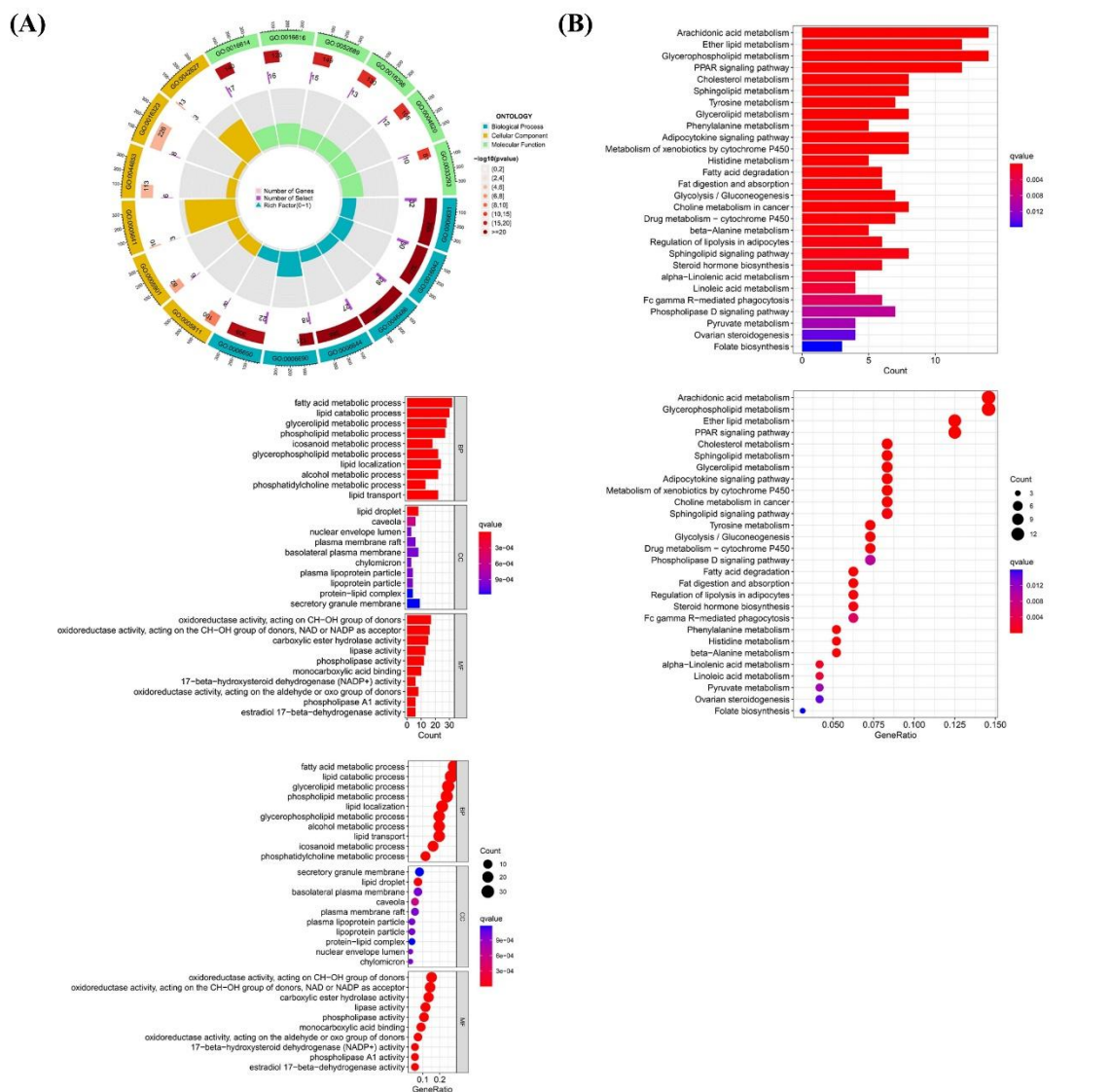


(A) Venn diagram showing the overlap between lipid metabolism-associated genes (LAGs) and genes represented in the TCGA-LUSC dataset, identifying the candidate LAG set used for downstream analysis.

1027 (B) Heatmap showing the expression patterns of lipid metabolism-associated
1028 differentially expressed genes (LADEGs) in LUSC and normal lung tissues from the
1029 TCGA cohort.

1030 (C) Volcano plot showing differentially expressed genes between LUSC and normal
1031 tissues, with highlighted LADEGs meeting the predefined significance criteria.

1032 **Figure 3. Functional enrichment analysis of lipid metabolism-associated**
1033 **differentially expressed genes in TCGA-LUSC.**



1034

1035 **(A)** Gene Ontology (GO) enrichment analysis of LADEGs, including biological process,
 1036 cellular component, and molecular function categories.

1037 **(B)** Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis of
 1038 LADEGs. The results indicate that the identified genes were mainly enriched in lipid-
 1039 related metabolic pathways and biologically relevant processes associated with tumor
 1040 metabolism.

Figure 4. Protein–protein interaction network analysis and hub-gene identification for LADEGs.

(A) Protein–protein interaction (PPI) network of LADEGs generated using STRING.

(B) Cytoscape-based visualization of the interaction network, showing the topological relationships among upregulated and downregulated genes.

(C) Subnetwork of the top hub genes identified from the PPI network. These genes were prioritized for downstream prognostic screening.

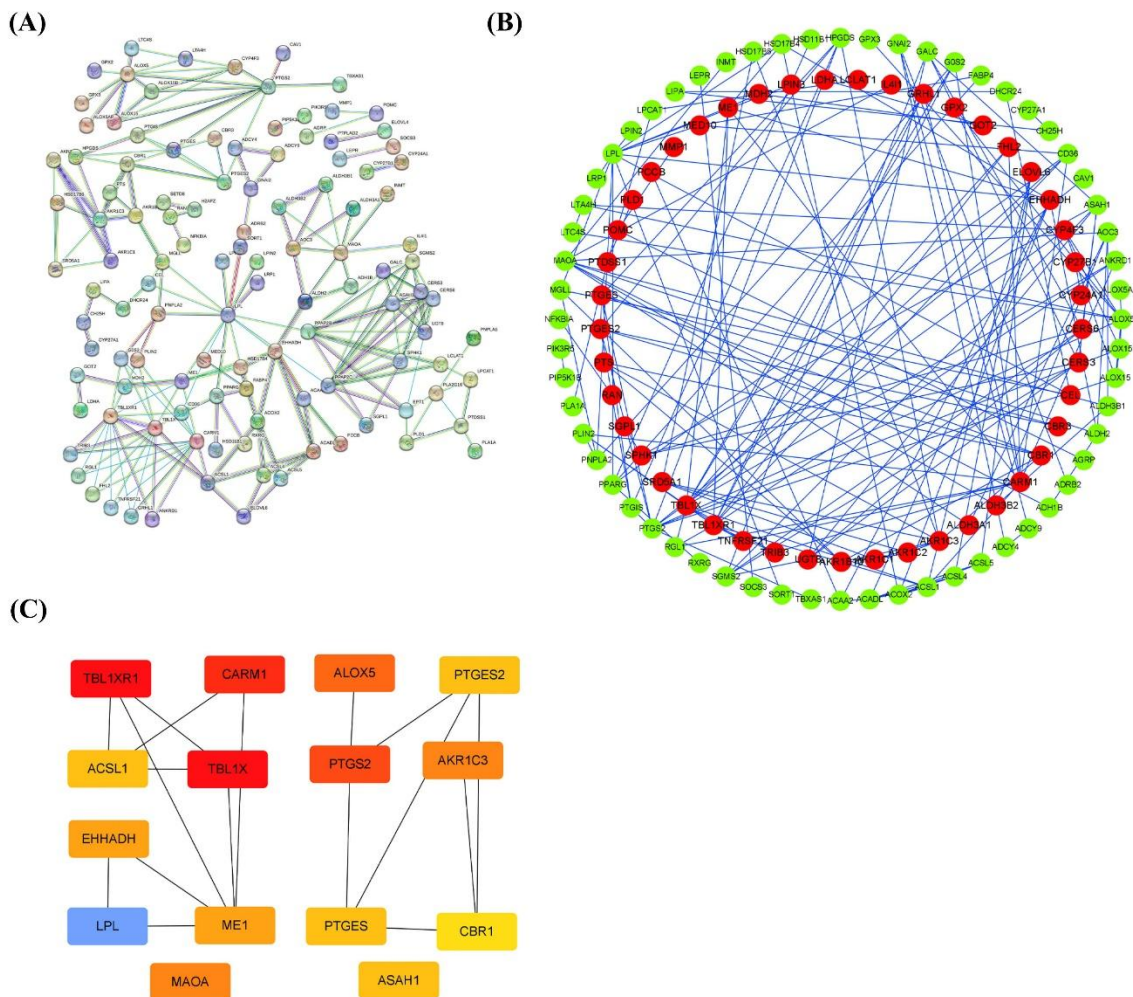
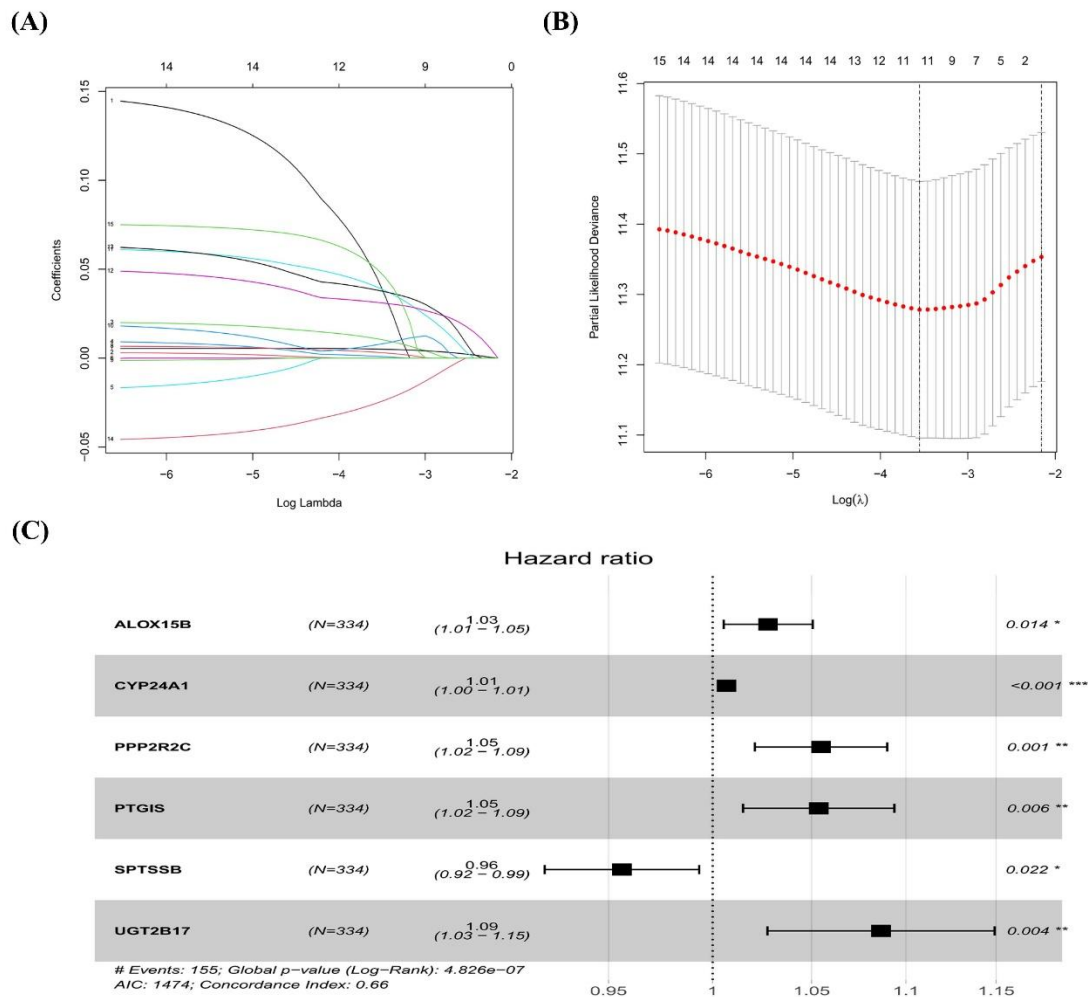


Figure 5. Construction of the six-gene lipid metabolism-associated prognostic signature in the TCGA training set.

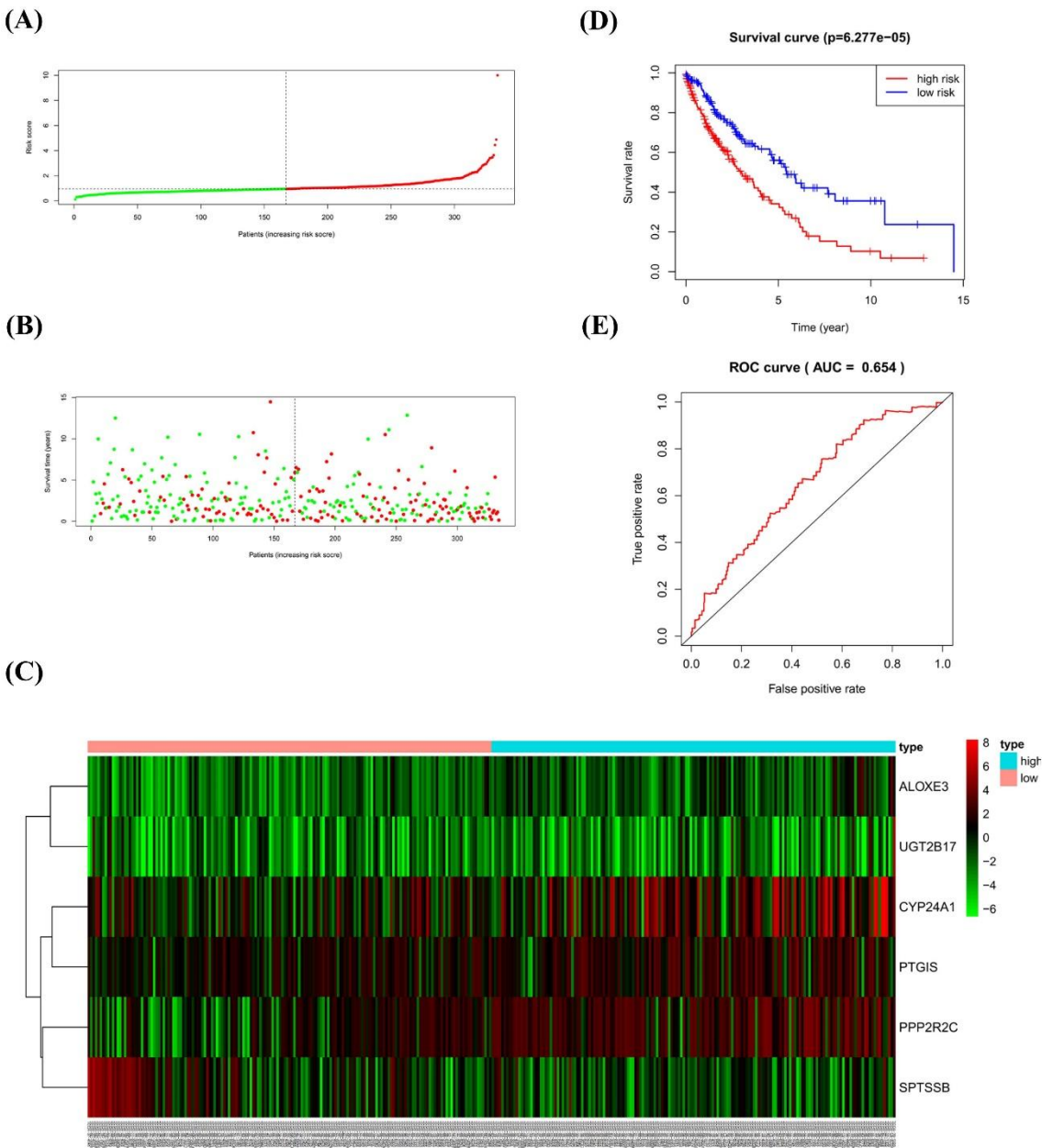


(A) LASSO coefficient profiles of candidate prognostic genes across the penalty parameter sequence.

(B) Cross-validation plot used to determine the optimal penalty parameter in the LASSO model.

(C) Multivariable Cox regression analysis of the final six genes included in the lipid metabolism-associated prognostic signature (LAPS).

Figure 6. Prognostic performance of the six-gene LAPS in the TCGA training set.



(A) Distribution of risk scores in the TCGA training cohort.

(B) Distribution of survival status and follow-up time according to increasing risk score.

(C) Heatmap showing the expression patterns of the six genes in the high-risk and low-risk groups.

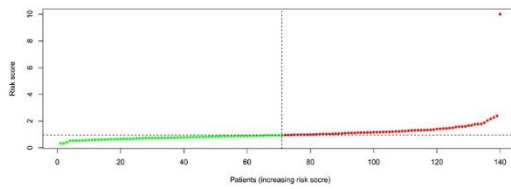
1065 (D) Kaplan–Meier survival curves showing overall survival in the high-risk and low-risk
1066 groups.

1067 (E) Time-dependent receiver operating characteristic (ROC) curve evaluating the
1068 discriminatory performance of the six-gene LAPS in the TCGA training cohort.

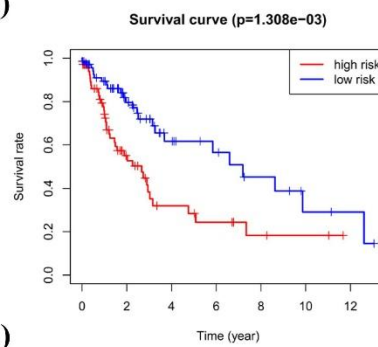
1069

1070 **Figure 7. Internal validation of the six-gene LAPS in the TCGA internal test set.**

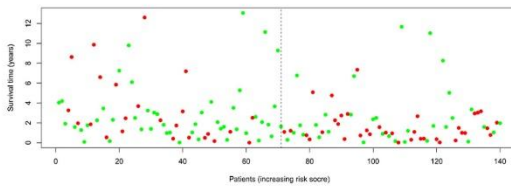
(A)



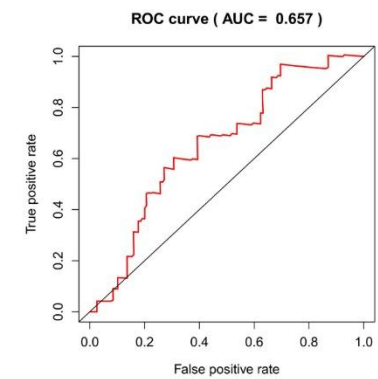
(D)



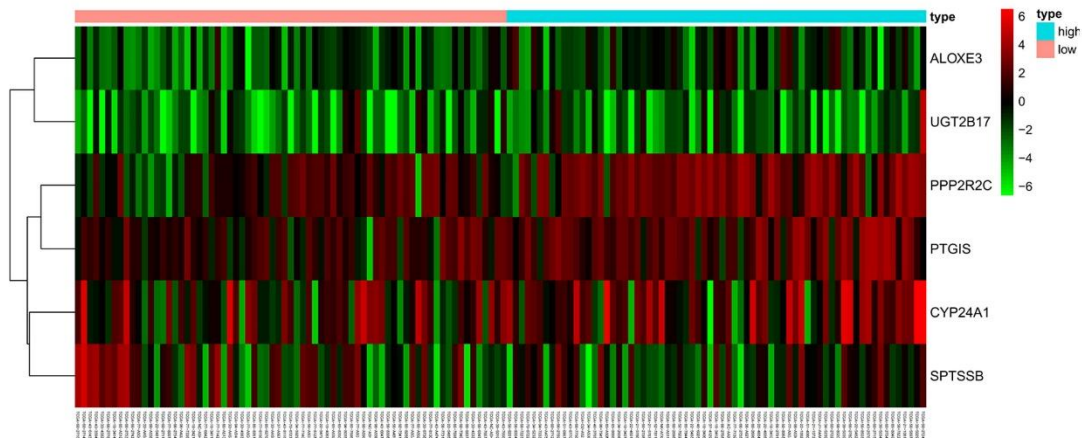
(B)



(E)



(C)



1071

1072 (A) Distribution of risk scores in the TCGA internal test cohort.

1073 (B) Distribution of survival status and follow-up time according to increasing risk score.

1074 (C) Heatmap showing the expression patterns of the six genes in the high-risk and low-risk groups.

1076 (D) Kaplan–Meier survival curves showing overall survival in the high-risk and low-risk

1077 groups.

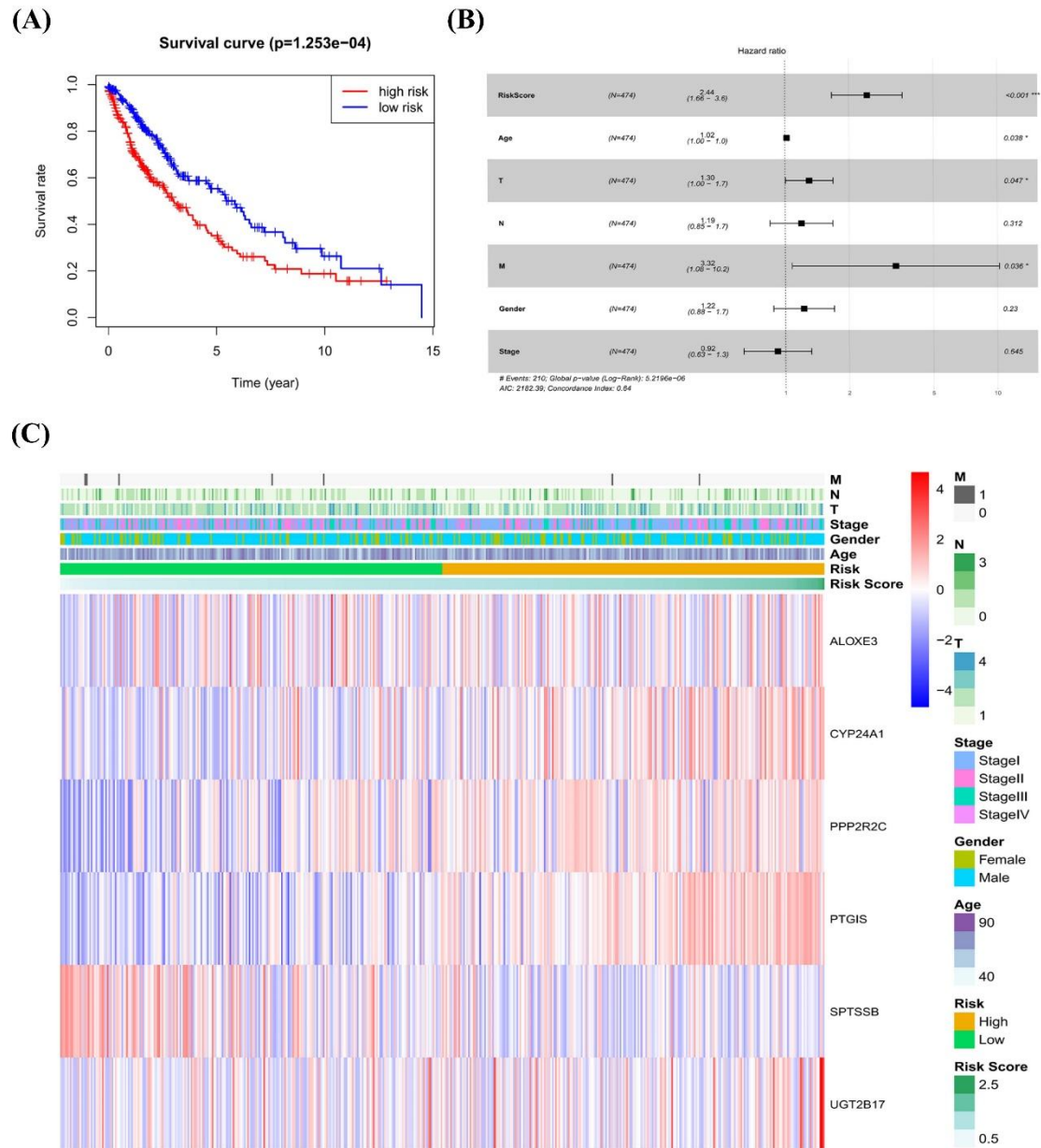
1078 **(E)** Time-dependent ROC curve evaluating the prognostic performance of the six-gene

1079 LAPS in the TCGA internal test cohort.

1080

1081 **Figure 8. Prognostic significance of the six-gene LAPS in the overall TCGA-LUSC**

1082 **cohort.**



1083

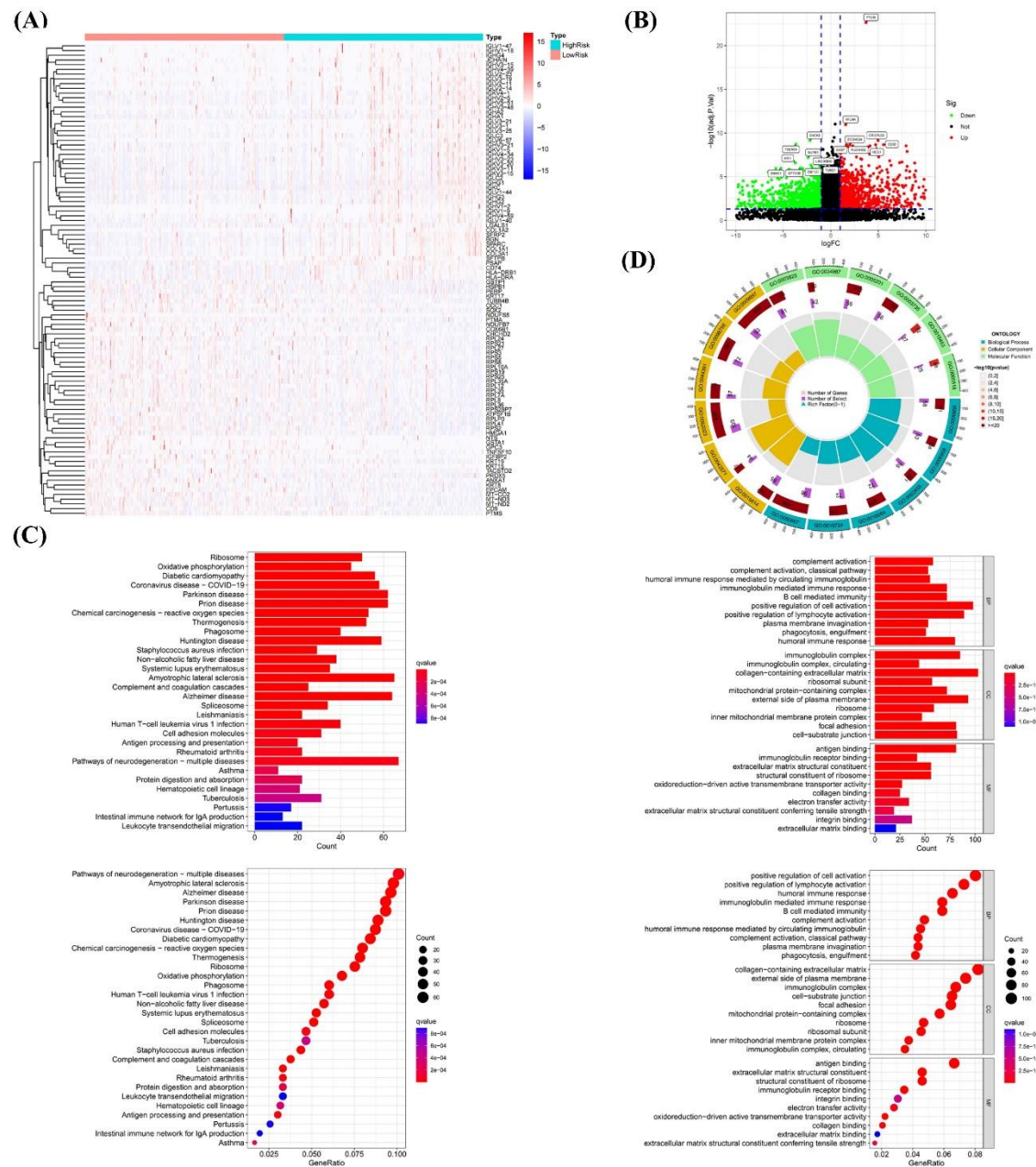
1084 **(A)** Kaplan–Meier survival analysis of the high-risk and low-risk groups in the overall
1085 TCGA-LUSC cohort.

1086 **(B)** Multivariable Cox regression analysis evaluating whether the risk score remained
1087 independently associated with overall survival after adjustment for clinicopathological
1088 variables.

1089 (C) Heatmap showing the distribution of the six-gene signature and corresponding
1090 clinicopathological annotations, including M stage, N stage, T stage, overall stage, sex,
1091 age, risk group, and risk score.

1092

1093 **Figure 9. Biological characteristics associated with the lipid metabolism-associated**
1094 **risk score in TCGA-LUSC.**



(A) Heatmap showing genes correlated with the LAPS-derived risk score.

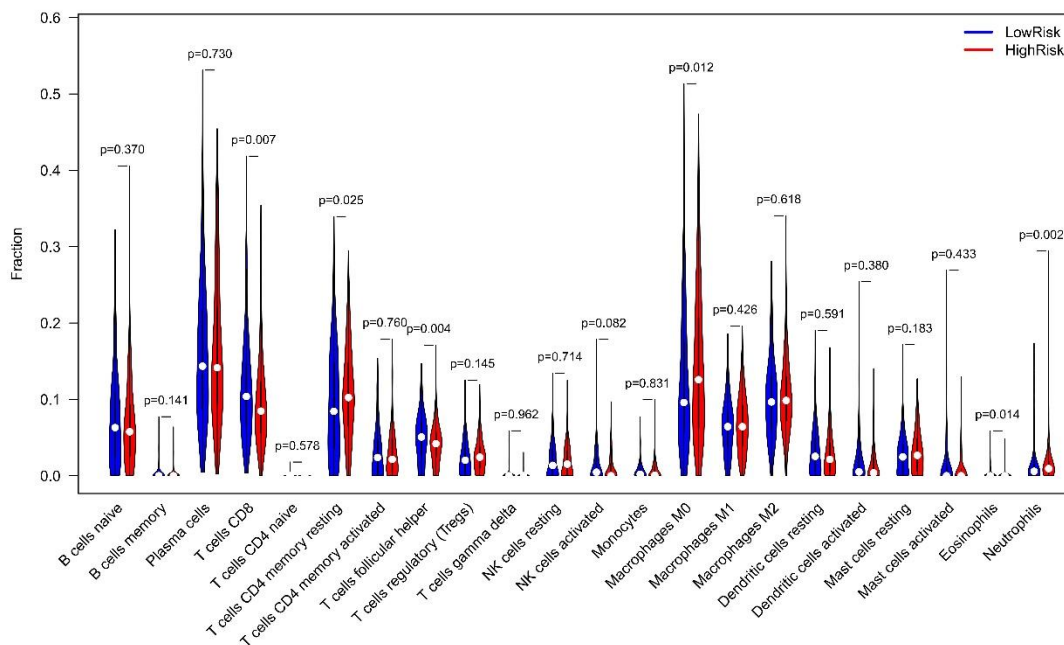
(B) Volcano plot showing genes positively and negatively associated with the risk score.

(C) KEGG enrichment analysis of genes associated with the risk score.

(D) GO enrichment analysis of genes associated with the risk score. These analyses

characterize biological pathways linked to risk-group stratification.

1102 **Figure 10. Immune-cell infiltration patterns in the high-risk and low-risk groups.**

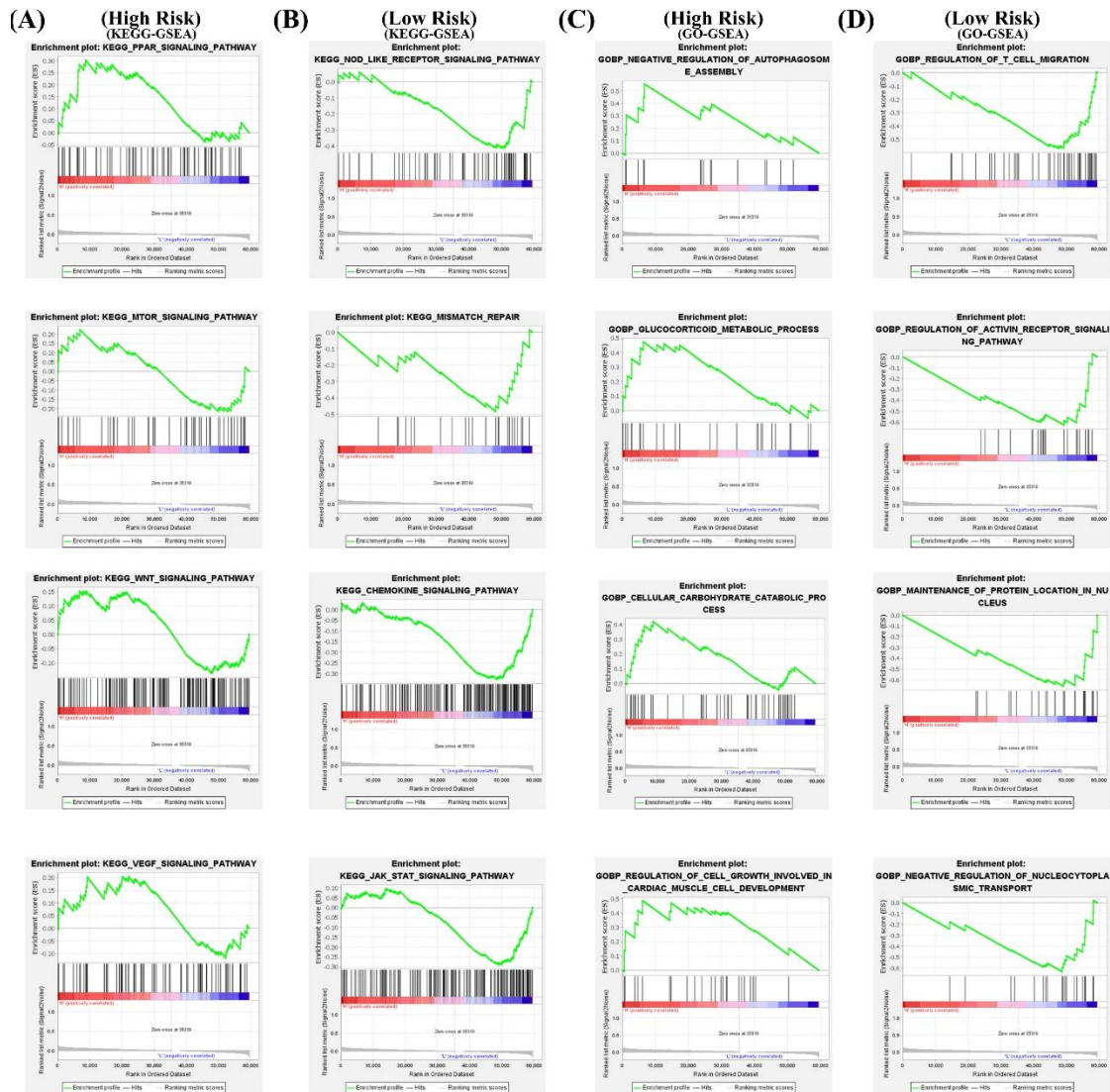


1103

1104 CIBERSORT-based comparison of tumor immune-cell infiltration between the high-risk
 1105 and low-risk groups in the TCGA-LUSC cohort. Violin plots show the relative fractions
 1106 of immune-cell subsets, and P values indicate between-group comparisons. These results
 1107 suggest distinct immune microenvironment characteristics associated with the lipid
 1108 metabolism-related risk groups.

1109

1110 **Figure 11. Gene set enrichment analysis of biological pathways associated with the**
 1111 **high-risk and low-risk groups.**

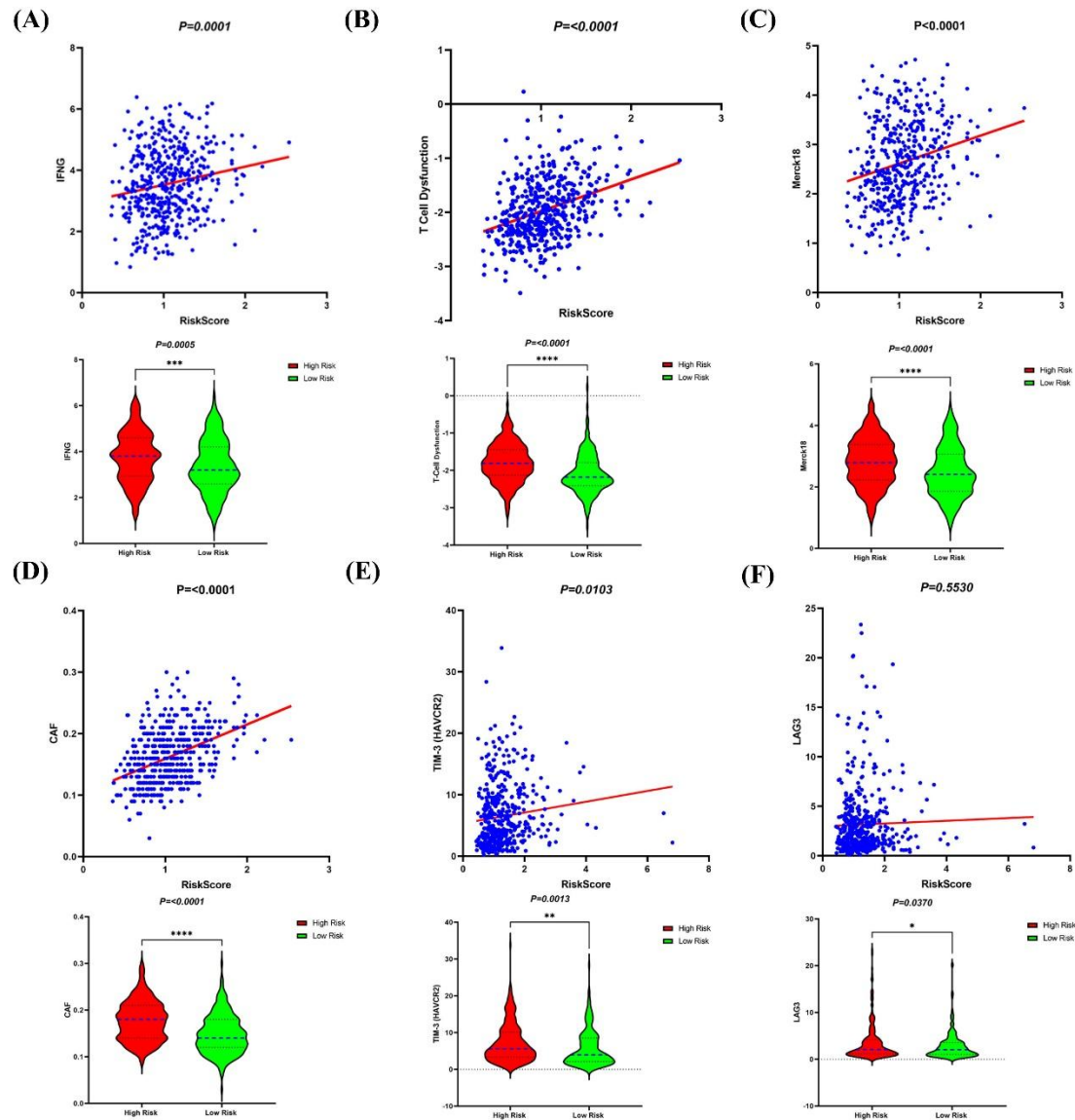


- 1112
- 1113 Representative GSEA enrichment plots for selected pathways in the high-risk and low-
- 1114 risk groups.
- 1115 (A) KEGG pathways enriched in the high-risk group.
- 1116 (B) KEGG pathways enriched in the low-risk group.
- 1117 (C) GO biological processes enriched in the high-risk group.
- 1118 (D) GO biological processes enriched in the low-risk group. These analyses highlight
- 1119 distinct pathway-level differences associated with LAPS-defined risk stratification.

1120

1121 **Figure 12. Exploratory associations between the six-gene LAPS and**

1122 **immunotherapy-response-related features in TCGA-LUSC.**



1123

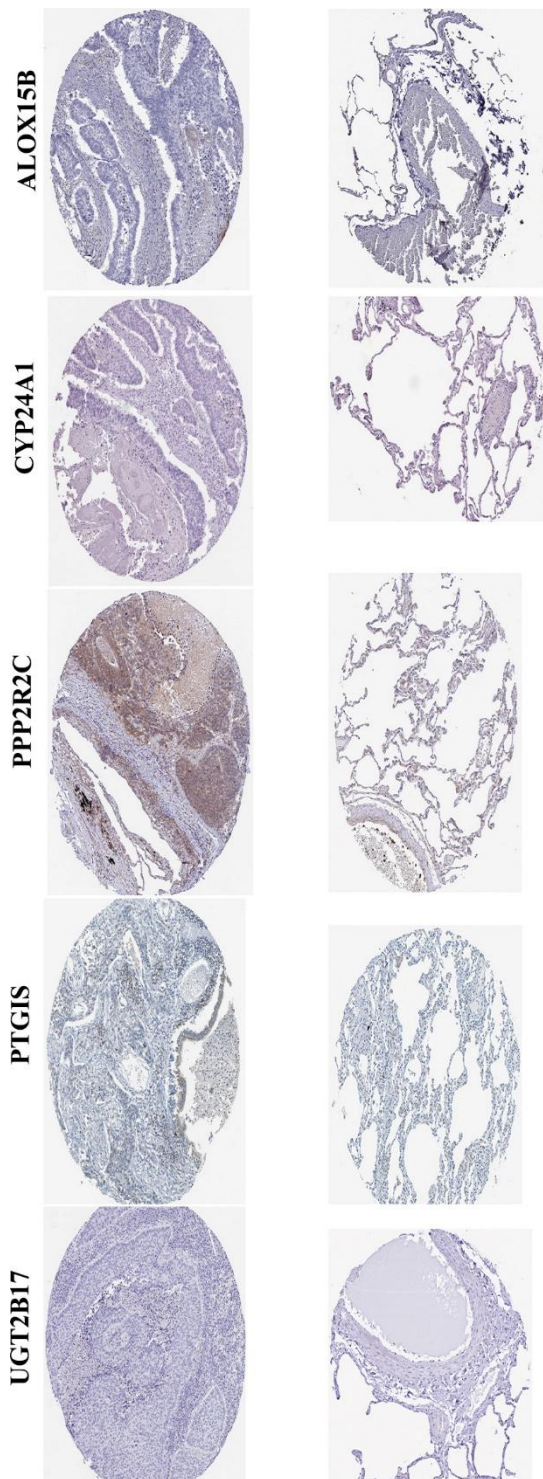
1124 **(A)** Correlation and group comparison between risk score and the IFNG score.

1125 **(B)** Correlation and group comparison between risk score and the T-cell dysfunction

1126 score.

1127 (C) Correlation and group comparison between risk score and the Merck18 score.
1128 (D) Correlation and group comparison between risk score and the cancer-associated
1129 fibroblast (CAF) score.
1130 (E) Correlation and group comparison between risk score and HAVCR2 (TIM-3)
1131 expression.
1132 (F) Correlation and group comparison between risk score and LAG3 expression. These
1133 analyses represent exploratory in silico associations relevant to immune contexture and
1134 immunotherapy-response-related features rather than direct clinical treatment-response
1135 validation.
1136

1137 **Figure 13. Public protein-expression support for selected genes within the six-gene**
1138 **signature in LUSC.**



1139

1140 Representative immunohistochemistry images from the Human Protein Atlas (HPA)
 1141 showing protein expression patterns of **ALOX15B**, **CYP24A1**, **PPP2R2C**, **PTGIS**, and
 1142 **UGT2B17** in LUSC and normal lung tissues. These data provide orthogonal public

1143 protein-level context for selected genes within the signature. HPA protein-expression
1144 data for **SPTSSB** were not available in the current dataset.

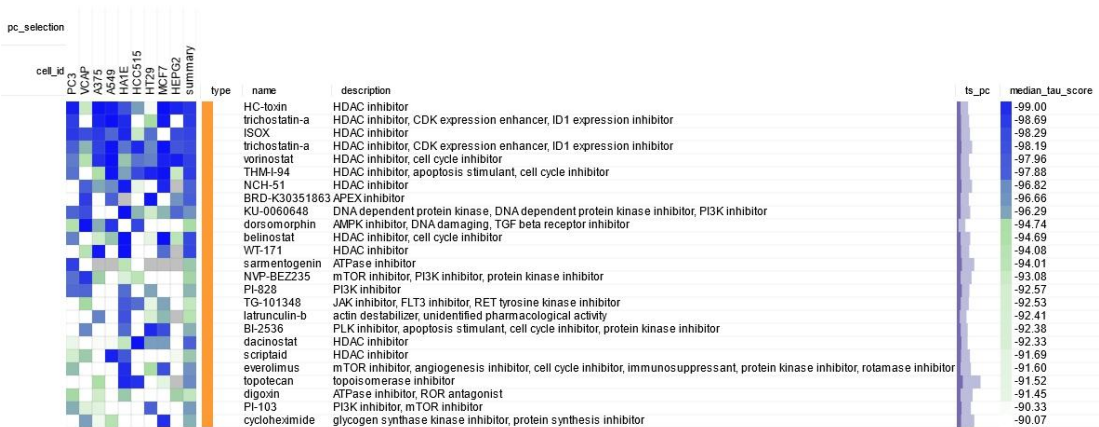
1145

1146 **Figure 14. Transcript-expression profiles of the six signature genes across lung**
1147 **cancer cell lines in the Human Protein Atlas.**

Public transcript-expression data from the Human Protein Atlas showing normalized transcript per million (nTPM) values for the six signature genes across lung cancer cell lines.

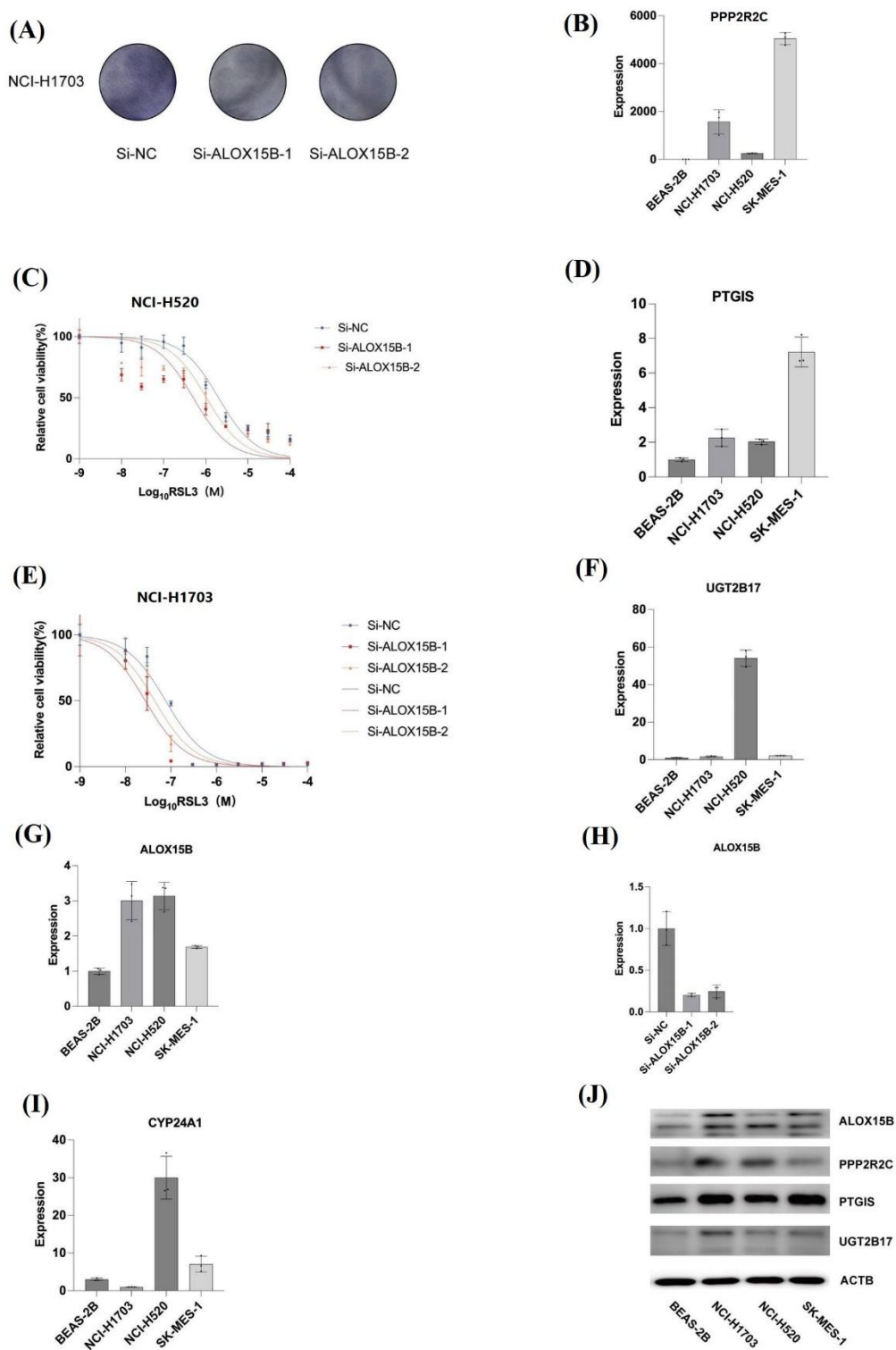
(A) ALOX15B, (B) CYP24A1, (C) PPP2R2C, (D) PTGIS, (E) UGT2B17, and (F) SPTSSB. These data provide public transcript-level context complementary to the in-house experimental validation.

Figure 15. Connectivity Map analysis identifying top-ranked candidate compounds associated with the LAPS-derived expression pattern.



Heatmap-style summary of top-ranked candidate small molecules identified using Connectivity Map (CMap) analysis. The figure shows perturbagen names, functional annotations, and connectivity-related scoring metrics, including touchstone percentile- and perturbagen-selection-related indices. These compounds represent hypothesis-generating candidates for future investigation rather than experimentally validated therapeutic agents.

1166 **Figure 16. Experimental evaluation of selected genes within the six-gene lipid**
1167 **metabolism-associated signature in LUSC cell lines.**



1168

Selected in vitro experiments were performed to provide biological support for genes within the prognostic signature.

(A) Representative colony-formation assay images of NCI-H1703 cells transfected with negative-control siRNA (Si-NC) or two independent **ALOX15B** siRNAs (Si-ALOX15B-1 and Si-ALOX15B-2), showing reduced clonogenic growth after **ALOX15B** knockdown.

(B, D, F, G, I) Baseline mRNA expression of **PPP2R2C (B)**, **PTGIS (D)**, **UGT2B17 (F)**, **ALOX15B (G)**, and **CYP24A1 (I)** in BEAS-2B, NCI-H1703, NCI-H520, and SK-MES-1 cells as determined by qPCR.

(C, E) RSL3-related cell-viability assays in NCI-H520 **(C)** and NCI-H1703 **(E)** cells after **ALOX15B** knockdown, showing altered viability responses across the tested concentration range.

(H) qPCR confirmation of **ALOX15B** knockdown after transfection with two independent siRNAs.

(J) Western blot analysis of ALOX15B, PPP2R2C, PTGIS, and UGT2B17 protein expression in BEAS-2B and LUSC cell lines; ACTB was used as the loading control.

Together, these experiments provide biological support for selected genes within the six-gene signature, with the strongest functional evidence centered on **ALOX15B**.

Tables

Table 1. Baseline clinicopathological characteristics of patients in TCGA-LUSC set.

1192

Baseline characteristics	Values*
Age (years)	62.5 (39-74)
Gender	
Male	372 (74.10)
Female	130 (25.89)
Stage	
I	244 (48.60)
II	162 (32.32)
III	85 (16.93)
IV	7 (1.39)
Unknown	4 (0.79)
T stage	
T1	113 (22.50)
T2	294 (58.56)
T3	71 (14.14)
T4	24 (4.78)
N stage	
N0	319 (63.54)
N1	132 (26.29)
N2	40 (7.96)
N3	5 (0.99)
Unknown	6 (1.19)

Metastasis	
Yes	7 (1.39)
No	412 (82.07)
Unknown	83(16.53)
Follow-up time (years)	1.81 (0-14.48)
EGFR mutation	
Yes	2 (0.39)
Unknown	500 (99.60)
KRAS mutation	
Yes	1 (0.19)
No	14 (2.78)
Unknown	487 (97.01)
Radiation therapy	
Yes	47 (9.36)
No	349 (69.52)
Unknown	106 (21.11)
Neoadjuvant treatment	
Yes	7 (1.39)
No	493 (98.20)
Unknown	2 (0.39)

1193 Abbreviations: LUSC: Lung Squamous Cell Carcinoma; TCGA: The Cancer Genome

1194 Atlas.* values are expressed as median (range) or n (%).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [suplementayfiles.rar](#)